Essays on Moral Hazard, Reputation and Market Structure

A thesis presented

by

Aron Toth

to

The Department of Economics

in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of

Economics

University of Warwick

September 2008

Contents

List of Tables

List of Figures

Acknowledgement

I have been incredibly fortunate to learn from and work with my supervisors. First and foremost, I am greatly indebted to Margaret Slade: without her continuous academic, personal and financial support this thesis would have never been accomplished. I also thank Amrita Dhillon who has shown distinct interest in my research from early on and whose support and encouragement was important throughout these years. Special thanks go to Claudio Mezzetti who took over my supervision at a difficult time. I am very grateful for the few but extremely insightful discussions; I learnt a lot in the short tenure of his supervision.

Throughout the years, I greatly benefited from the discussions with two friends and colleagues, Matthew Dickson and Andrea Salvatori. I also thank Chuck Blackorby, Bhaskar Dutta and Jeremy Smith for their special attention. The financial support from the Department of Economics at the University of Warwick is gratefully acknowledged.

Finally, I would also like to thank various seminar participants at the University of Warwick (2006, 2008), the Econometric Society Meeting (2006 Chennai), the Royal Economic Society Annual Conference (2007), the European Association for Research in Industrial Economics Annual Conferences (2007, 2008) and at the Bank of England (2008) for helpful comments and suggestions.

Declaration

I hereby declare that I have not used before or published any material contained in this thesis, it is my own work, and it has not been submitted for a degree at another university.

All Chapters of this thesis are single authored.

Summary

This thesis is comprised of three pieces of research on moral hazard, reputation and market structure. In particular, following an opening discussion of previous literature, I explore the dynamic interaction between moral hazard and market structure in two distinct game theoretic settings and empirically test a fundamental assumption of these models concerning consumer rationality.

In the first Chapter, I survey the studies which shed light on some dimension of the relationship between asymmetric information and market structure and identify the gap in the literature that my research aims to fill. The mechanism of reputation has been primarily investigated in the setting of perfect competition; however, this setting is ill suited for uncovering the rich set of relations between asymmetric information and market structure. Only a handful of articles departed from the perfect competition framework and only few of those introduced strategic interaction among firms, a fundamental ingredient of my research interest. The models which do include strategic interaction have, however, ignored some important dynamics in the interaction of asymmetric information and market structure.

Therefore in Chapter II, I develop a model in which market structure affects moral hazard while, in turn, moral hazard fuels market structure dynamics. The model is very general allowing for all kinds of strategic interaction among firms usually considered in the literature. I identify and analyse an important driving force - a survival contest - which has so far been overlooked. The main conclusion is that market concentration in and of itself reduces moral hazard and moral hazard drives the market towards concentration through the survival contest. The model is suitable to explain the puzzling market transformation of important industries such as banking, audit and health care.

In Chapter III, I extend the model of Chapter II by introducing stochastic entry. First, I demonstrate that my results in the previous Chapter are robust to the entry process. Second, stochastic entry allows me to derive a non-degenerate steady state distribution which exhibits a very intuitive dynamics. Finally, although the complex nature of the dynamics prevents a detailed comparative static analysis of this distribution, it displays two well known empirical regularities. In particular, my model shows that the presence of moral hazard in and of itself produces shake-outs in the market from time to time and also correlated exit and entry rates.

The reputation mechanisms in general and in the models of Chapter II and III in particular crucially depend on consumers' ability and willingness to develop an understanding of imperfect information on quality. In order to make reputation an effective disciplinary force, consumers must be strongly rational so that they read and understand imperfect quality indicators. In Chapter IV, this basic assumption on consumer rationality is tested empirically in discrete choice settings in the audit market. I find robust empirical evidence that if consumers are firms rather than individuals, they are strongly rational.

Chapter I

On the Interaction of Reputation and Market Structure

1 Introduction

Asymmetric information is pervasive in many markets. Namely, it is often the case that one of the self-interested parties has more information than the other and also a natural incentive to exploit this informational advantage. This can jeopardise mutually beneficial cooperation between the parties. Repeated interactions and the threat of losing long term gains from cooperation are often sufficient to prevent opportunistic behaviour. The mechanism through which the informational problem can be tackled and long term cooperation can be achieved is often coined as reputation. Reputation models have been in the forefront of applied research for more than two decades, since they offer an effective market mechanism through which asymmetric information can be overcome. Reputation models are also of interest in the theory of repeated games in general. The predictive power of repeated games is often very limited because typically any mutually beneficial outcome can be sustained as an equilibrium (folk theorem). Reputation models provide a useful way to select among the multiplicity of equilibria in these games. (Kandori 2006)

The aim of this Chapter is to review studies which are related to the interaction of reputation and market structure. The literature on reputation is vast and therefore, the review does not endeavour to be exhaustive in any way. Rather, I focus on the important insights of some of the seminal papers. Reputation has been discussed primarily in two settings: unobserved quality and labour market (principal-agent) applications. Here, I only discuss articles within the unobserved quality literature. The reasons are twofold. First, I am primarily interested in the dynamic interaction of reputation and market structure, and the latter is rarely incorporated in labour applications. Second, the main insights are quite similar in the two literatures. I will also not discuss many other interesting models of reputation (e.g. markets for reputation), which go beyond the well defined theme of reputation and market structure.

First, I briefly review articles which have established fundamental concepts and uncovered the basic driving forces behind reputation building and maintenance. Most of this literature discusses models in the framework of perfect competition. The principal question of these analyses is to what extent perfect competition is a threat to quality in the presence of asymmetric information. These articles, therefore, investigate only one particular market structure and its effect on dynamic incentives. Since these studies ignore the strategic interaction among firms, they cannot give a general description of the dynamic relations between reputation and market structure. Second, I review articles which depart from the perfect competition setting in one way or another and attempt to uncover some particular dimensions of the dynamic interaction between reputation and market structure.

Before turning to the survey, let me briefly review some basic concepts and terminology related to reputation and repeated games.

2 Basic concepts

Throughout this Chapter, the problem of asymmetric information and reputation will be discussed in the framework of unobserved quality. In particular, there are firms wishing to produce and sell a good which can be of either high or low quality. While producing the good, firms can exert effort or "invest in quality" (action) that typically yields a high quality product (outcome). There are consumers wishing to buy the good and they prefer high to low quality. Consumers, however, cannot observe the investment (or effort) since it is private information. In addition, consumers are unable to observe product quality (outcome) before purchase; they only observe it ex post, after consumption.

Reputation models are often categorised along the nature of asymmetric information. One class of games is the adverse selection or hidden information models. In these games, the buyer is uncertain about key aspects (i.e. "the type") of the firm, which do not depend on the firm's action, but only on its (hidden) information. Reputation here is a result of learning: buyers may infer the type of the firm through her past performance. Basically, these games show how rational buyers try to weed out the bad types through their evolving beliefs, how they learn to overcome informational asymmetry and distinguish between good and bad firms.

Another class of games is the moral hazard or hidden action models. In these models, there is only one type of a firm, but some firm's action is not observed by consumers. These models are not so much about the learning process but more about the buyer trying to find the right set of (dynamic) incentives which induce firms to produce the desired quality. In other words, through the aid of strategic threats and punishments, buyers try to effectively deter the firm to abuse her informational advantage. Lastly, there are mixed models that are of both moral hazard and adverse selection.

In adverse selection models, reputation usually builds up over time as buyers learn about firms through past performance. In these models the game often converges to its complete information counterpart over time. In particular, either the buyers learn the type of the firm eventually, or the different firm types play identical strategies.

In moral hazard models, however, reputation is often a result of an equilibrium in which the firm never chooses to cheat. In other words, in adverse selection models the focus is usually on "reputation build-up", reputation acquisition and evolution, whereas in moral hazard games the central issue is more about "reputation maintenance". Note that, as opposed to adverse selection, in moral hazard models the length of the horizon is often a central issue. The reason is that in equilibrium firms choose to not cheat as long as the discounted present value of this behaviour is bigger than the value of deviation. Therefore, in order to make reputation maintenance attractive enough, an infinite horizon with a sufficiently large discount factor is usually necessary.

A second dimension along which the literature is differentiated is the type of monitoring (perfect versus imperfect and public versus private). Perfect monitoring occurs when although firms' actions are not observed ex ante, they are observed perfectly ex post. So, for instance, if the firm's high effort results in high quality with certainty, then from the quality outcome observed after consumption the buyers can deduce what action the firm has taken. On the contrary, if high effort leads to high quality only with some probability strictly less than one, than the observed outcome (quality) is only a signal of the firm's action, hence monitoring is imperfect. As we will see shortly, early studies of reputation predominantly assumed perfect monitoring, while later models were mostly cast in the framework of imperfect monitoring. Imperfect monitoring is more general and makes it typically more difficult to acquire and maintain a reputation.

When the (perfect or imperfect) signal is publicly observed, the model is one of public monitoring; otherwise, it is one of private monitoring. I do not discuss models with private monitoring because they are not related to the theme I am interested in. However, I briefly mention that these models are usually quite complicated because there is no common knowledge of histories and hence punishments which are conditional on these commonly observed histories are not feasible. As a result strategy coordination can break down. (e.g. punishment is a best response if everyone else punishes, etc.)

3 Reputation and perfect competition

The unobserved quality literature is often thought to have started with the seminal work of Klein and Leffler (1981). In this paper, the effects of reputation is analysed in a competitive market in which firms do not act strategically among themselves. This article lays out the basic concept behind reputation, namely how the prospect of future profits can discipline opportunistic behaviour by firms. In a model of moral hazard and perfect public monitoring, Klein and Leffler demonstrate that when quality is unobserved before purchase the equilibrium price must be above average cost otherwise the firm will feel inclined to cheat and sell a low quality product at a high quality price. In particular, their main result was to show that even in dynamic settings competitive pricing cannot be reconciled with high quality provision. The insight is simple: consumers' maximum punishment strategy, that is, their threat of ceasing to buy from the cheating firm cannot possibly induce high quality production because future profits are zero anyway with competitive prices. There must be a price premium in order to induce high quality provision. A price premium, however, would attract endless entry. Their solution is to introduce some firm-specific capital investment which is equal to the discounted present value of the future profit stream in equilibrium. In other words, the paper uses a dynamic concept of competition: the total present value of the firm is zero. Note that this firm specific investment would not occur if quality were observable. That is, it is strictly speaking unnecessary and socially wasteful: the utility consumers get from these investments does not justify its costs otherwise they would be undertaken anyway. Many of successor papers aim to address this unappealing feature of the Klein and Leffler model. (e.g. Shapiro 1983, Allen 1984)

Shapiro (1983) provides a more formal structure for many of the insights in Klein and Leffler (1981) in a bit more general framework (heterogeneous consumers, range of quality levels) and supplements it with welfare analysis. However, the perfectly competitive environment (the absence of strategic interaction) and moral hazard with perfect public monitoring remain. In this model, the firm's reputation is perceived by consumers to be its last period quality. His analysis differs from Klein and Leffler (1981) in that the "fixed cost of entry" is endogenous and comes from the fact that an entrant must establish its reputation. That is, in the first period the entrant is able to sell only at low price due to lack of reputation. It produces high quality, however, because the next period consumers will know her last period quality and from then on she can charge the appropriate high quality price. This investment in reputation in the first period, similarly to Klein and Leffler (1981), results in a zero discounted present value of profits, but without any socially wasteful firm-specific capital investment. The unappealing feature of this solution is that consumer beliefs are not rational in equilibrium, since they always expect low quality from an entrant while she always produces high quality. Shapiro (1983) also explores his baseline model under the assumption that consumers adjust their beliefs adaptively, and hence reputation evolves with some lag so the firm's misconduct feeds into consumer beliefs only gradually.

Diamond (1989) investigates reputational dynamics in a model of adverse selection and moral hazard. Although he presents his analysis in the context of debt markets, it may be instructive to discuss his findings in the unobserved quality framework. Diamond introduces the concept of imperfect public monitoring into the reputation model and this is one of the distinctive features of his model. In particular, while in the previous studies actions were not observed ex ante, they were observed ex post. That is, if a firm chose to invest little in quality (chose a low cost technology), it did produce low quality and the market learnt its misconduct ex post. In Diamond's model, however, a firm produces low quality only with a certain probability, which is a function of its investment. In other words, the firm's actions are directly observed neither ex ante, nor ex post. As a result, consumers can deduce the firm's actions in equilibrium from the quality outcomes only over a longer period, similarly to Shapiro's extended model with adaptive consumer beliefs. This naturally increases the profits from deviating to low effort and hence increases the incentive to cheat. Consequently, despite consumers' unforgiving behaviour (maximum punishment), in Diamond's model the equilibrium is not stationary and hence imperfect public monitoring leads to more complex reputation dynamics. In particular, if the adverse selection is severe (there is a large fraction of firms able to produce only bad quality), then the market expects a low quality initially on average and therefore the price will be low. At the start, this low price is not sufficient to give the necessary incentives to firms who have the potential to produce high quality to do so (moral hazard). On the other hand, some firms will produce high quality even if their investment was low as a result of the assumption that quality is a stochastic function of investment, rather than deterministic. To these surviving firms, the market is willing to pay higher price in subsequent periods since their track record distinguishes them. Therefore, those firms which survive invest more and if sufficient reputation is built up (the price is high enough), then firms will invest in high quality from then on.

Hörner (2002) has many similarities with some of the previous models. His study is also a model of both moral hazard and adverse selection with imperfect public monitoring. The main objective of his article is to demonstrate that the continuous exertion of high effort in equilibrium can indeed be compatible with perfect competition and in this sense it improves on Diamond's results. Hörner uses the idea of dynamic competition introduced by Klein and Leffler (1981) and Shapiro (1983). That is, in his model both inept (adverse selection) and normal firms (moral hazard) make zero profits over their expected lifetime. The intuition is as follows. In equilibrium, consumers abandon a firm if it has ever produced bad quality (maximum punishment strategy). Knowing the initial share of inept firms, the probability of high quality at high and low effort and their own strategy, in equilibrium consumers can calculate the expected number of normal firms in the market, that is, the expected quality. Since more inept firms than normal firms fail in each period, inept firms disappear slowly and hence the expected quality increases over time. Therefore, evolving consumer beliefs support a monotonically increasing price schedule in equilibrium. This increasing price profile is the incentive that induces normal firms to exert high effort throughout their life. In particular, normal firms do not cheat because then they live longer enjoying higher prices towards the end of their expected lifetime. Bad firms make more profit at the beginning since they don't exert costly effort but they also live less giving them zero expected discounted profit stream in total. On the contrary, normal firms exert effort and realise less profit at the beginning but benefit from higher prices later since they stay on longer. Both firm types realise zero expected profit overall.

The spirit of the reputational dynamics in Hörner (2002) is very similar to that of Shapiro (1983): in Shapiro's model negative profits at the initial stage were necessary to compensate for the price premium later on. Hörner eliminates the unappealing feature of Shapiro's model; in Hörner's model consumer expectations are rational. His article is novel in one more important respect. Reputational effects typically weaken over time for two reasons. First, if the horizon is finite, then approaching the end of the game, the continuation payoff shrinks giving less dynamic incentive to maintain reputation. Second, with imperfect public monitoring as beliefs evolve and consumers become convinced that the firm exerts high effort, producing high quality improves on customers' belief less and less, so the marginal benefit of maintaining high reputation decreases. In addition, if consumers are sufficiently convinced that the firm does not cheat and low quality is just the result of bad luck, they will not abandon the firm. But then, the firm has an obvious incentive to cheat. These properties together or separately typically produce non-monotonous reputation dynamics in the long run. (Bar-Isaac and Tadelis 2008) However, as Hörner shows in his study unforgiving consumers can maintain high quality provision indefinitely as long as there is competition so they have the option to switch.

4 Reputation and imperfect competition

The studies so far have typically investigated to what extent high quality provision can be reconciled with perfect competition. The articles discussed below depart from the perfect competition framework. This strand of the literature focuses more on the question that I analyse in subsequent Chapters.

Allen (1984) recasts the model of Klein and Leffler by introducing strategic interaction among firms in the form of Bertrand competition. The model remains one with moral hazard and perfect public monitoring. The basic insight of the study is that if consumers are sophisticated enough and know the technology, observe prices as well as quantities, then they can pin down firms' behaviour and deter firms from cheating even with competitive prices. Allen observes that there is no need for wasteful firm-specific investments in order for a competitive high quality equilibrium to exist. First-best high quality provision is possible in competitive environments. However, depending on the cost structure, first-best equilibrium may not exist; equilibrium may be second-best, with incumbents that do not produce at the minimum of average costs – the situation presented by Klein and Leffler. However, even Allen's second best equilibrium results in lower prices than in the Klein and Leffler model, since wasteful firm-specific investments are not present in his model. The basic driving force behind Allen's findings is sophisticated and super-rational consumers whose maximum punishment strategy have a supreme bite. His consumers are more sophisticated than that of Klein and Leffler since they know the technology and observe quantities in addition to prices. As a result, they can pin down firm behaviour much more effectively. Allen (1984) argues that in fact his main results hold even if quantity is not observed, because quantity can be inferred from price in equilibrium. This is because Allen assumes Bertrand competition, and not, as Klein and Leffler, that firms are price takers. In Allen's model, due to strategic interaction, a firm cannot expand production unless it lowers its price, while in the model of Klein and Leffler firms can expand production infinitely at the going price. Thus, in Allen's model, whether the product is of high quality can be detected ex ante even if quantities are not directly observed, because any deviation of a profitable output change (unobservable) must be associated with some price move (observable).

Kranton (2003) introduces a model of moral hazard with perfect public monitoring where firms compete for market shares via prices. She makes the important assumption that consumers do not link prices and quality. That is, prices do not serve as signals for quality. She proves that there are circumstances under which there exists no (constant) price which would support a subgame perfect equilibrium in which all firms produce high quality. The intuition is simply that competition can eliminate the price premium needed to sustain high quality production if consumers fail to link prices and quality.

Bar-Isaac (2005) studies a model somewhat similar to Kranton (2003). His framework is of moral hazard with perfect public monitoring. In his analysis, the level of competition is exogenous and quantity (rather than prices) is the strategic device of firms in addition to effort (investment in quality). He also makes Kranton's crucial assumption of consumers not taking quantity as a signal of quality. That is, consumers evaluate only past quality performances in order to form their beliefs about current quality. He finds a non-monotonic effect of competition on quality. This is the result of the following effects. On the one hand, competition decreases the profits in high quality equilibrium in an obvious way and, hence, the incentive to invest in quality. On the other hand, it can also affect the cost of losing reputation. Once a firm produces bad quality it acquires a bad reputation forever in equilibrium. The cost of losing reputation is essentially the difference between the values of a firm with good and bad reputation. This cost may increase with competition if the value of bad reputation is not invariant to the level of competition. Note that in Kranton's model bad reputation always resulted in a stream of zero profits and hence the value of the firm with bad reputation was unaffected by the level of competition. Through instantaneous profits, competition, on the other hand, naturally reduces the value of the firm with good reputation so the difference between good and bad reputation firm values undoubtedly decreases with competition in her model. In Bar-Isaac (2005), the firm value of bad reputation is not zero and it is very much affected by competition, just like the value of the firm with good reputation. These two equilibrium firm values can change with competition at different rates. As a consequence, the loss of reputation may result in a bigger fall in market share in more competitive environments and this increases the cost of losing reputation and hence decreases the incentive to deviate to low effort. Therefore, while competition has a direct and undoubtedly negative effect on the incentive to invest in quality through the reduction of current profits, it can also have an indirect positive effect through increasing the cost of losing reputation. Which one dominates depends on market structure: in his model highly competitive and highly concentrated markets could sustain high quality in equilibrium but not intermediate market structures.

Dana and Fong (2008) address the same question, that is, under what conditions strategic interaction among firms is a threat to high quality equilibrium. The framework is also the same; it is a model of moral hazard with perfect public monitoring and the level of competition is exogenous. They also focus on the effects identified by Bar-Isaac, namely, how the (exogenous) level of competition changes the cost and benefit of deviations when cheating does not necessarily imply a stream of zero profits. In other words, just like in Bar-Isaac (2005), low quality good is of value to consumers and can be produced profitably. One of the new features is that while Kranton and Bar-Isaac assume that consumers do not take prices (quantity) as a signal for quality, Dana and Fong relax this assumption. This can have quite a dramatic effect on the results of previous papers. They find a non-monotonic relationship between market structure and quality too but they reach exactly the opposite conclusion to that of Bar-Isaac (2005). In their model it is easier to sustain high quality in oligopoly than in monopoly or competitive markets. They construct an equilibrium in which consumers link quality and price in a specific way. In particular, in oligopoly if a firm deviates from the high quality price, consumers expect low quality and, as a consequence, firms revert to low-quality and marginal cost pricing from then on. That is, a change in price (or quality) triggers a price war. This makes the punishment particularly severe and hence easier to sustain high quality provision in an oligopoly. In contrast, in the case of a monopoly, the firm would still be able to charge a (lower) monopoly price once its reputation is lost so the difference between the discounted values of good and bad reputation are not that big after all making it more tempting to deviate. Similarly, in a competitive environment the difference between the equilibrium values of the good and bad reputations is relatively small because although the present value of a firm with bad reputation is zero under perfect competition, the value of good reputation isn't very big either. In sum, by deviating the monopoly loses little because it can still make substantial profits without reputation. By deviation, a firm in fierce competition loses little too because it didn't have much to lose in the first place. A cheating oligopolist, on the other hand, loses much more because a firm with good reputation is relatively profitable while the situation of a firm with bad reputation is just as bad as under perfect competition.

Rob and Fishman (2005) develop a model of reputation with moral hazard and imperfect public monitoring in a market of local monopolists. As discussed above, reputation is usually a stationary phenomenon in models of moral hazard. Rob and Fishman (2005) introduce dynamics through non-stationary equilibrium actions, similarly to Diamond (1989). In particular, local monopolists do not always exert high effort: younger firms tend to shirk while older firms are more disciplined. The basic idea behind this dynamics is that older monopolists have a bigger customer base, more profitable and hence they have more to lose. As a consequence, they invest more in quality and fail less often. Therefore, in equilibrium the industry is dominated by bigger firms along with smaller ones constantly trying their luck. The customer base expansion is a result of the assumption that information about quality flows among customers through word of mouth. That is, if in one period the firm produces good quality, then a mass of new consumers proportional to the current customer base will learn about it and patronise the firm in the next period. The good news of high quality has a bigger impact on future demand for bigger firms. A distinctive feature of this model is that while in some models like Isaac-Bar (2005) and Kranton (2003) consumers are assumed to not link prices with quality, Rob and Fishman construct the equilibrium in a way that prices are not signals for quality. In other words, in their model there is no signalling role for prices in equilibrium; this is essential in order to introduce dynamics in reputation models with moral hazard. I will follow this modelling approach in Chapter II and III.

5 Conclusion

In sum, high quality provision may be reconciled with competitive prices under very special circumstances. In all cases, however, consumers must be very sophisticated and have a complete knowledge of the technology. In addition, strategic interaction among firms can be a threat to high quality provision with less sophisticated consumers.

However, the effect of strategic interaction, and hence the dynamic relation between market structure and reputation has only been partially explored. In particular, all of the models with strategic interaction among firms focus only on one aspect of the relationship between moral hazard and market structure. That is, they consider models where competition and market structure naturally influences profits earned and, therefore, the incentive to invest in quality. However, these papers ignore the other equally important aspect of the relationship: moral hazard can affect market structure too, which in turn can affect the strategic behaviour of firms. In other words, market structure is not endogenous, only a parameter in these models. However, if consumers can force firms to fail, then it seems reasonable that firms will recognise that they are in a game of gambling for survival. Firms must ask themselves the question: what do I gain if others fail? Providing answers to this question is one of the goals of Chapter II and III.

Chapter II

The Great Industry Gamble: Market Structure Dynamics with Moral Hazard

In this Chapter, I investigate the dynamic effect of moral hazard on market structure in a general framework. In this model the evolution of market structure determines the severity of moral hazard and, in turn, moral hazard fuels market structure dynamics through a survival contest. In the absence of scale economies I show that the presence of moral hazard results in a convergence towards market concentration regardless of the intensity of competition. On the other hand, the dynamics leading to market concentration reduces moral hazard even when prices do not increase with concentration (e.g. Bertrand competition). Therefore, the main policy implication is that market concentration can be effective against moral hazard and as such, welfare increasing. The model is suitable to explain the puzzling market transformation of industries such as banking, health care and audit.

1 Introduction

Consumers often cannot verify certain key product characteristics before consumption (experience and credence goods).[1] The firm, therefore, has a natural incentive to exploit its informational advantage and cheat on consumers - this is a classical example of moral hazard. Interestingly, many markets where the unobserved product characteristics (quality) are crucial have a tendency towards concentration over time despite the fact that scale economies are often not substantial. Moreover, there is a common perception that concentration has a positive effect on quality in these industries even when the intensity of price competition does not ease with concentration. These tendencies are particularly prominent in previously regulated markets where competition of some sort has been introduced. For instance, in banking, health care and audit, concentration has been increasing steadily since market liberalization while price competition appears to be fiercer than ever. Traditional arguments (e.g. scale economies), however, often fail to offer a credible explanation for these peculiar industry dynamics.

I aim to rationalise these market characteristics in a simple infinitely repeated oligopoly model emphasising the key role of moral hazard. The dynamic relationship between moral hazard and market structure is explored and fully characterised in the absence of scale economies. I argue that market evolution determines the severity of moral hazard while, in turn, moral hazard drives market structure through a survival contest. I show that the presence of this basic driving force does not depend on the nature of price competition. In particular, a firm can invest in (unobserved) quality and this investment increases its probability of survival through the reduction of moral hazard. The contest is a result of strategic gambling on survival: firms try to outlive rivals so that they can be among the few (or only one) to benefit from lucrative market structures. However, more rivals upset the prospect of this survival battle, hence firms care little about the future. Consequently, they invest less, thereby exacerbating the moral hazard problem and reducing quality. This triggers a shake-out when the market is shared by many. Failing firms, on the other hand, raise the stakes for survival since the market gets closer to favourable states. Therefore, surviving firms invest more and hence alleviate the moral hazard problem even when prices do not increase with concentration (e.g. Bertrand competition). Counterintuitively, this means that a firm in a more concentrated market can be *less* profitable along the equilibrium path. Increasing the chances of survival, remaining firms fail with smaller and smaller probability slowing down the convergence towards concentration and turning the contest into a ruthless struggle. Failure rates decrease with concentration as firms find it more and more difficult to outlast rivals. This in turn forcibly reduces the chance that

[1]These key product characteristics are usually labelled as quality in the literature and hereafter I will follow this convention. I would like to stress that what I consider quality here is the set of *key* characteristics in a sense that they essentially determine the value of the product. Quality, therefore is different from "amenities", which are rather designed to give the appearance of quality. So for example, in a hospital, quality would be the effectiveness of clinical treatment (unobserved), whereas waiting times, complaint resolutions, etc (observed) would be less crucial characteristics.

the market will ever reach the lucrative states and leads to markets along the equilibrium path which are shared by only a few firms and yet characterised by cut-throat price competition, very little moral hazard and relatively high quality. The effect of the intensity of price competition is secondary to this survival contest. Lax competition induces more investment reducing moral hazard and as such supports a less concentrated market in the long run. Intense rivalry in prices, on the other hand, speeds up the convergence towards concentration and results in fewer firms in equilibrium.

Naturally, the relationship between moral hazard and market structure has important policy implications. This paper shows that in markets where moral hazard is especially severe, market concentration can alleviate moral hazard. Interestingly, the more intense the competition, the more important it is not to block the natural evolution of market structure and to allow for mergers and restricted entry.

1.1 Stylized facts and empirical evidence

The industry dynamics described above can be observed in markets which were previously regulated. In particular, once regulatory restrictions are lifted and competition is unleashed, the market goes through a rapid transformation. The initial phase is characterised by poor quality and heavy shake-out in the form of failures, mergers and acquisitions. Yet, as the market is driven towards concentration, failure rates decrease while quality often improves substantially even when cut-throat price competition appears to remain. Conventional arguments for an industry transformation of this sort (e.g. scale economies) do not always seem to apply. Three industries of this kind have received particular attention in the last couple of decades: banking, health care and audit.

A. The Banking Industry

In the banking industry, I consider quality to be the risk that a bank chooses to take by the design of its investment portfolio. This risk (the probability of insolvency) is never directly observed by the depositors (moral hazard). However, once the bank's financial position is in doubt, depositors withdraw their money and the bank often fails.[2]

From the 80's the deregulation and liberalization process of the financial systems started to gather momentum worldwide. As a result, competition started to challenge banks which led to a failure, merger and acquisition wave of an unexpected magnitude almost all over the world. The most striking example

[2] One may ask: why would the depositor care about risk of insolvency if deposits are insured? First, deposit insurance is never designed to give full protection for standard moral hazard reasons. Second, it's not only the deposit that a client can lose in the case of bankruptcy. With time, consumers build up valuable relationship with the bank, which gets lost if the bank goes bankrupt. In this particular industry, failure can also be thought of as an event independent of consumer's attitude towards risk (quality); that is, in banking it doesn't really matter if consumers care about risk or not since banks can fail as a direct result of their investment decisions.

is the US where the number of banks has decreased by 40% since 1984 but market concentration has also increased steadily in almost all developed and developing countries in the last couple of decades. (Rare exceptions are Finland and France) This spectacular concentration process has long been a puzzle in the academic literature since empirical research has more often than not failed to find significant economies of scale. (Bikker 2004) Furthermore, it is well documented in numerous empirical studies that, as a direct consequence of the liberalization process, banks started to engage in riskier activities. (Keeley, 1999; Allen and Gale 2000; Brewer III and Jackson III 2006) However, as banking markets steadily tended towards a concentrated structure, less and less risk was observed to be taken. (Maudos and Guevara 2004, Beck et al, 2005)[3] Beck et al (2006), for instance, demonstrated on a sample of 69 countries over the period of 1980-1997 that more concentrated banking markets are less prone to systemic banking crises.[4]

As a consequence, this perceived negative relationship between risk and market concentration (the "concentration-stability" paradigm) has widely been appreciated by competition authorities: despite the apparent lack of efficiency gains from mergers, antitrust policy has been particularly lenient in the financial industry in the last couple of decades. (Boyd and Graham 1996, Boyd and Nicolo 2005)[5]

B. The Health Care Industry

Moral hazard in health care is very apparent: a patient is almost never in the position to observe and verify the quality of clinical treatment she receives. In fact, measuring quality of health care is a notoriously difficult task even for researchers. Hence, like in banking, for decades a puzzling policy question in health care has been whether the total welfare effect of competition is positive or negative.

While competition undoubtedly reduces health care costs in the long run, its effect on quality is ambiguous. The controversies of quality measurement in health care are particularly severe since the quality of clinical treatment is very poorly observed. It is a difficult task to find hard empirical evidence on the effect of competition on quality. Not surprisingly, the empirical findings on the relationship between competition and quality is very mixed and largely dependent on the way quality is defined and measured. (Wong 2004, Pauly 2004, Romano and Mutter 2004) To my knowledge, there is only one article which avoids the problem of measuring quality explicitly and yet convincingly succeeds in giving

[3]Risk taken by financial institutions is of utmost importance because bank failures could easily lead to systemic banking crises, which usually have tremendous social costs. In the USA, for example, between 1984 and 1991 more than 1400 savings and loans company and 1300 banks failed, resulting in clean up costs of 3.2% of the GDP. In many cases the direct cost of systemic crises have been around 10-20% of the GDP, with occassional magnitude of 40-55% of GDP (Chile, Argentina). (Caprio et al, 1996)

[4]In their article, a systemic banking crisis is defined by emergency measures taken to assist the banking system, and/or non-performing assets that reached 10% of the total assets, and/or fiscal rescue operations that exceeded 2% of the GDP. It may also be worthwhile noting that in their sample 47 crises are included.

[5]I should emphasize that the theoretical and empirical evidence on the negative relationship between risk and concentration is best described as mixed. (e.g. Beck et al 2006) This is largely due to the fact that financial risk is very complex and can be measured in numerous ways. However, the very fact that merger policy appears to have adopted the "concentration-stability" view suggest that this relationship may indeed be prevalent.

some important insight into the relationship of market concentration and quality. Jin (2005) investigates health maintenance organizations' (HMO) voluntary disclosure of quality. The market of HMOs is highly competitive. Since the early 1990s, these organizations have been widely criticized for the presumed low service quality, and the market observed many failures as well as mergers and acquisitions. (Jin 2005) As a response, an independent and non-profit agency, the National Committee of Quality Assurance (NCQA) started to accredit these institutions on a voluntary basis. By 1998 around half of the HMOs were accredited. Jin (2005) finds that the proportion of HMOs which chose to voluntarily disclose information on quality via accreditation is significantly higher in more concentrated markets. This certainly appears to be at odds with theory as well as common sense unless less concentrated markets indeed provide lower quality.[6]

Similarly to banking, recent empirical studies provide only minimal evidence of scale economies in managed care: such economies have been found to exist only at relatively low levels. (Glied 2000) The findings from these articles are in line with merger studies because mergers have minor effects on health care costs. (Glied 2000) As a consequence of an ongoing consolidation in the HMO market, 79% of MSAs were highly concentrated (HHI>3000) by 2004. (AMA 2005)[7]

C. The Audit Industry

The market for audit services is another typical example of credence goods. The certification of the accuracy of financial statements by a third party is based on a credence claim: the integrity (quality) of an audit is never directly observed. The independent auditor faces a dilemma. On the one hand, if it is more lenient about the accounting standards, more willing to bend the truth, more companies will want to subscribe to its audit services. On the other hand, the more false the picture the auditor agrees to certify (that is, the more corrupt the auditor), the more probable it is that the actual performance of its clients reveals the auditor's misconduct (for example, one of the audited firms unexpectedly fails), and hence the higher the threat of litigation and reputational damages and, as a consequence, of failure. Audits are primarily produced for shareholders and potential investors who rely on the credibility of the audit and ultimately, the credibility of the auditor. Once credibility is seriously questioned, the product (audit) is worthless and the auditor quite possibly goes bankrupt. Essentially, this is the stylised story of Arthur Andersen's failure in 2002. By the time of its indictment for misconduct, Arthur Andersen had already lost nearly all of its clients.

[6] In theory, in competitive markets informational asymmetries can be overcome by certain market mechanisms such as voluntary quality disclosure via independent quality verification agencies. A high quality producer has a natural incentive to disclose information on its quality through these agencies, distinguishing itself from other firms which produce inferior quality. For further details see the references in Jin (2005).

[7] For broader (product) market definition of combined HMO and Preferred Provider Organization (PPO) market, the corresponding percentage would be 67%.

Despite fierce rivalry, we observe an extremely concentrated audit market today although this was not always the case. In the USA, for instance, the audit industry came into existence as a result of the Securities Acts of 1933 and 1934. At the beginning, competition among auditors was severely impaired: advertising, competitive bidding and soliciting clients were all strictly prohibited. As a result, hundreds of audit firms were operating undisturbed for three decades. Beginning from the 70s, however, rules restricting auditors from advertising and competitive bidding were loosened unleashing fierce competition. Ever since, the industry has been going through a remarkable concentration process. By the 1980s, eight firms dominated the American audit market and by 1998, this was down to five as a result of a series of mergers in 1980s and 1990s among the big 8. Then, the collapse of Arthur Andersen further reduced the number of dominant firms to four. Whereas only 22% of the smaller companies (revenue less than $100 million) were audited by the big 4, the large company market is very concentrated: 96% of the companies with revenues over $500 million were big 4 clients in 2006. (GAO-08-163) Unfortunately, unlike in the banking and health care markets, data on audit costs are not publicly available, therefore, direct tests for economies of scale in the industry are simply hard to come by. However, after a careful inspection of the predominantly labour intensive audit technology, it's difficult to imagine why substantial economies of scale should exist in this industry that could credibly determine such a high level of concentration.

Recent events such as the bankruptcy of Arthur Andersen raise similar policy issues to banking and health care: is concentration good or bad, does a concentrated market promote quality or, on the contrary, does it actually harm the integrity of audit? (Bar-Yosef and Sarath 2005)

In this paper, I build a model which aims to rationalise the following stylized facts: i) Steady convergence towards concentration in the absence of scale economies, ii) High failure rates in fragmented markets, and iii) Product quality increasing with market concentration. The model shows that instead of technological reasons, the presence of moral hazard can be the key driving force behind the puzzling market evolution observed in banking, health care and audit industries.

1.2 Related literature

In the context of unobserved quality, asymmetric information has been investigated extensively under static competitive conditions. For instance, Allen (1984) presents a model of moral hazard where highly sophisticated consumers infer quality from prices and an equilibrium exists where firms always produce high quality in Bertrand competition. Hörner (2002) shows that high quality can be sustained in perfect competition under adverse selection and moral hazard. More recently, Rob and Fishman (2005) argue that local monopolists provide higher quality as their size grows when information about quality flows among consumers by word of mouth. However, only Allen (1984) introduces strategic interaction among firms and none of these articles addresses market structure dynamics explicitly. In addition, Kranton (2003), Bar-Isaac (2005) and Dana and Fong (2008) consider models where competition and market structure influence profits earned and, therefore, the incentive to invest in quality. Their main concern is how and to what extent different competitive conditions can undermine the incentive to invest in quality. In these papers the dynamics of market structure is not analysed since they focus on one direction of the relationship between market structure and moral hazard: market structure affects moral hazard. These papers ignore the other important aspect of the relationship: moral hazard can affect market structure and, in turn, the strategic behaviour of firms. In particular, if consumers can prompt failure, then it seems reasonable to think that firms recognise that they are in a gambling game for survival and hence they must ask: what do I gain if others fail? This is the driving force that the current study focuses on.

In sum, my paper differs from previous studies because it explicitly models the full extent of the dynamic relationship between market structure and investment in quality (moral hazard) while allowing for all of the possible forms of strategic interactions among firms considered in the literature. The paper identifies a key driving force that has been overlooked so far: a survival contest.

The paper is organized as follows. In Section 2, I discuss the main assumptions and outline the basic model with and without entry. I discuss the main results in Section 3 and conclude in Section 4. Appendix A contains an extension of the model, which explicitly incorporates consumers into the game discussed in the main body. Finally, lengthy proofs are contained in Appendix B.

2 The model

Denote the number of firms in the market by $n \in \{0, 1, 2, ...\}$. Each individual firm $i = 1, ..., n$ produces a (homogenous or heterogenous) good, the quality of which can be of two types, high or low. Neither consumers nor firms observe quality before consumption; it is commonly observed after consumption. Consumers obtain zero (positive) utility from a low (high) quality good.[8] Firms can invest in quality (or exert effort) in every period and this investment is private information (moral hazard). The probability of producing high quality in a given period depends only on how much the firm invests in that period. There is no knowledge accumulation or increase in efficiency as a result of past investments which could potentially introduce scale economies. Firm i can invest in quality $x_i \in [0, 1]$ at the cost of $g(x_i)$ each period and produces high quality at the end of that period with probability x_i (i.e. imperfect public monitoring). I assume that a firm that produces low quality fails and leaves the market.

Assumption 1. *If a firm produces low quality, it fails.*

[8] The good, therefore, can be heterogenous only along observable characteristics.

Assumption 1 is purely technical and can be relaxed by incorporating consumers explicitly into the game and showing that if they ever experience low quality, consumers rationally stop buying from that firm in equilibrium. The details of this extended game can be found in Appendix A. This is very standard consumer behaviour in the unobserved quality literature (e.g. Allen 1984, Hörner 2002). There is, however, one crucial difference between the optimal consumer strategy in the models of Allen (1984) and Hörner (2002) and the consumer strategy in the current game. In their models, the equilibrium configuration is such that either consumers are aware of industry specifics so they can calculate the price that ensures high quality (Allen 1984) or they simply hold the "right beliefs" about the lowest price necessary for high quality (Hörner 2002). In other words, price has a signalling role in these models. This requires high level of consumer sophistication so in my modelling approach I follow Rob and Fishman (2005) and construct an equilibrium in which consumers' conduct does not link quality and price.

As a consequence of Assumption 1, a firm fails with probability $1 - x_i$ at the end of the period and hence in any given period firms that have produced only good quality are in the market. Therefore, including firm i, the number of firms (the market structure) at the beginning of the next period will be $n - k$ with probability $x_i \Pr(k|x_{-i})$, where $x_{-i} = [x_j]_{j \neq i}$ and $k = 0, ..., n-1$ is the number of firms other than firm i failing at the end of the period. $\Pr(k|x_{-i})$ is the probability mass function of the convolution of $n - 1$ Bernoulli distributions with "success" probabilities x_{-i}.

At the beginning of each period, firms engage in price competition. For simplicity, I do not explicitly model this price game. Denote the symmetric equilibrium profit for a firm from the price game by $\pi(n)$, when n firms are in the market.

Assumption 2. *$\pi(n) \leq \pi(n-1)$, for $n > 2$ and $\pi(2) < \pi(1)$ with $\pi(\cdot) \geq 0$.*

The per period gross profit is weakly decreasing in the number of firms, non-negative for oligopoly and the per period gross monopoly profit is strictly positive. Importantly, note that Assumption 2 does not impose any restriction on the nature of price competition. It is general enough to account for homogeneous-product as well as differentiated Bertrand or Cournot competition. This specification essentially captures all the cases typically considered in the literature.

Having realised per period gross profits, firms invest in quality. Hence, the per period net profit for firm i is $\pi(n) - g(x_i)$. I make the following assumptions on the cost of investment:

Assumption 3. *$g(x_i), g'(x_i), g''(x_i) > 0$ for $x_i > 0$ and $g(0) = g'(0) = 0$, $g(1) < \infty$.*

The cost of investment is a strictly increasing convex function. The assumption of bounded investment costs is not essential, the main results are unaffected if I relax it. Its only purpose is to ensure that a corner solution ("high quality equilibrium") may exist so that I can compare my results to the findings of previous studies in a straightforward manner. As we will see, without entry, this can also be a way to ensure that the market structure converges to a steady state. Were the cost to tend to infinity, $\lim_{x_i \to 1} g(x_i) = \infty$, then not even a monopoly would find it optimal to invest the maximum amount ($x_i = 1$) with finite gross profits. As a consequence, even a monopoly would fail sooner or later and the market would vanish in the long run. With entry, however, we do not need the assumption of bounded costs in order for the market structure to converge to a steady state, as we will see in Section 2.2.

In most markets $g(x_i)$ has a natural interpretation. In the banking industry $g(x_i)$ can be understood as an opportunity cost: the lower risk (higher quality) the bank takes, the more potential profit it sacrifices. The price game in banking, therefore, could be thought of as competition for deposits. Suppose there are investment opportunities for banks in every period ranked by their risk. Having acquired deposits at the equilibrium interest rate in the stage game, the bank can invest in the riskiest project available and can ensure a profit of $\pi(n)$ in that period. However, investing in less risky assets, the bank willingly decreases its potential profit in order to increase the probability of survival into the next period and acquires a per period net payoff $\pi(n) - g(x_i)$. This net profit can even be negative if the bank offers loans at a lower rate than it acquires deposits.[9]

Assumption 4. *There is no entry.*

This assumption is maintained in Section 2.1 for expositional purposes. However, I will relax it in Section 2.2.

Time is discrete and infinite. The common discount factor is $\beta \in (0, 1]$. I consider the quality game in the framework of an infinitely repeated game.[10] The summary of timing in this (reduced) game is as follows:

1. Given n, gross profits, $\pi(n)$, are realised.
2. Firms simultaneously choose investment in quality, x.
3. Failures and exits occur.

The strategies are assumed to be Markov and I focus on symmetric Markov Perfect Equilibrium (MPE) throughout the analysis. That is, strategies depend only on payoff relevant information. The

[9] Bounded costs in Assumption 3, therefore, are not unreasonable. In the banking industry, for instance, if the risk of the available investment opportunities is bounded, there is certainly a maximum amount that a bank can sacrifice in order to secure survival unless one is willing to entertain the idea of negative credit rates.

[10] Since the results on moral hazard in the reputation literature are often sensitive to the length of the horizon, perhaps I should emphasise that the findings of the current model readily go through with finite horizon as well; in fact, two periods would be sufficient.

payoff relevant information can be conveniently "condensed" into a state variable, which is the number of firms n. For expositional purposes, I first present the model without entry.

2.1 The game without entry

Assumption 1-4. are maintained throughout this section. With n firms in the market, the dynamic programme of firm i is simply

$$v(n;x_{-i}) = \pi(n) + \max_{0\leq x_i \leq 1}\left\{-g(x_i) + \beta x_i \sum_{k=0}^{n-1} V(n-k)\Pr(k|x_{-i})\right\} \tag{1}$$

where n is the current state and $V(n-k)$ is the value of the firm in a symmetric equilibrium at state $n-k$. Specifying the (endogenous) transition probabilities explicitly, we get

$$\begin{aligned} v(n;x_{-i}) &= \pi(n) + \max_{0\leq x_i \leq 1}\Big\{-g(x_i) + \beta x_i\Big[V(n)\prod\nolimits_{-i} x_{-i} + \\ &\quad +V(n-1)\sum\nolimits_{l=1,l\neq i}^{n-1}(1-x_l)\prod\nolimits_{-l} x_{-l} + \ldots + V(1)\prod\nolimits_{-i}(1-x_{-i})\Big]\Big\} \end{aligned}$$

After differentiating and imposing symmetry, the first order conditions (FOC) of this programme are

$$\beta\sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^k x^{n-k-1} - g'(x) + \lambda - \mu = 0, \qquad \lambda x = 0, \qquad \mu(x-1) = 0 \tag{2}$$

where λ, μ are Lagrange multipliers. The simple nature of the first order conditions is the result of two facts. First, the convolution of Bernoulli trials is simply the binomial distribution due to symmetry. Second, the Bernoulli probabilities are linear. The latter may appear overly simplistic. However, given the non-linear costs, it is not. In other words, it is not more restrictive than assuming non-linear survival probabilities but constant marginal costs, a commonly applied modelling structure (e.g. Rob and Fisherman 2007, Ericson and Pakes 1995). In the current model the marginal benefit is constant (given rivals' strategy) and the marginal cost is not, whilst in the other models it's exactly the opposite. In both cases the trade off between a unit cost and a unit benefit is non-linear and the first order conditions yield non-linear strategies in the state variable.

Denote the solution to the first order conditions (2) by $x(n)$. Given Assumption 2 and 3, it is easy to see that $x(n) \neq 0$. There are two possible solutions: one corner and one interior. That is, $x(n) = 1$ as long as

$$\pi(n) \geq g'(1)\frac{1-\beta}{\beta} + g(1) \tag{3}$$

Alternatively, $0 < x(n) < 1$ where $x(n)$ solves (2) with $\lambda = \mu = 0$. Before we proceed, let's have a closer look at the corner solution which represents the high quality equilibrium, the traditional focus of the literature on reputation. If (3) holds, then firms will invest the maximum amount in quality and there will be no failures at the end of the period and the market structure will not change. The necessary condition for this to happen is that firms generate enough revenue from the price game to cover the cost of producing the highest quality. If they don't discount the future, that is, if they are infinitely patient ($\beta = 1$), then non-negative net profits ($\pi(n) - g(1) \geq 0$) are sufficient for the market structure being at a standstill. However, the more impatient the firms, the higher the net profit they must make, the higher the price premium the consumers must pay for the highest quality. In the limit, as $\beta \rightarrow 0$, in the static game, no firm would ever produce high quality, regardless of profits. In other words, as impatience of the firms grows, the incentive is bigger to cheat on consumers since firms care less about survival. Inequality (3) is consistent with many of the previous findings in the literature on reputation. (e.g. Allen 1984)

I first establish the existence of a symmetric pure strategy equilibrium.

Proposition 1 *In the game without entry, there exists a symmetric Markov Perfect Equilibrium in pure strategies.*

Proof. See Appendix B. ∎

The next Proposition follows from the first order conditions directly:

Proposition 2 *In a symmetric equilibrium of the game without entry, in the long run the number of firms, n^*, is no bigger than $\tilde{n}$ ($n^* \leq \tilde{n}$), where*

$$\tilde{n} = \begin{cases} 0 \text{ if } \pi(1) < g'(1)\frac{1-\beta}{\beta} + g(1) \\ \max\{n : \pi(n) \geq g'(1)\frac{1-\beta}{\beta} + g(1)\} \text{ otherwise} \end{cases}$$

It is immediate from (3) that the higher the cost of investment in quality, the smaller the number of firms viable in the long run. Also, perhaps paradoxically, more intense competition leads to more concentrated markets since depressed gross profits discourage investment in quality. Greater patience, however, as captured by the discount factor, supports the existence of more firms in a long run equilibrium.

The following Proposition states the main result. It shows that equilibrium investment rises with concentration even if gross profits $\pi(n)$ don't. As a consequence, paradoxically a firm with only a few rivals may realise *less* profit along the equilibrium path than a firm with many competitors.

Proposition 3 *In a symmetric equilibrium without entry, more firms in the market implies strictly less investment in quality, that is,* $x(n+1) < x(n) < 1$, *unless* $n+1 \leq \tilde{n}$, *in which case* $x(n+1) = x(n) = 1$.

Proof. See Appendix B. ■

Proposition 3 states that, unless gross profits are sufficiently high, $n+1$ firms will always invest strictly less than n firms. However, if gross profits in a market with $n+1$ firms are big enough (inequality (3) holds), then firms in this market (and any other market with fewer firms) will invest the maximum amount. In the "high quality equilibrium" therefore, no firm fails and the market never gets more concentrated. In other words, if the market structure is such that $n > \tilde{n}$, then firms will fail with positive probability ($x(n) < 1$) and the market steadily converges towards $\tilde{n}$. Therefore, $n^* \leq \tilde{n}$ are the only viable market structures in the long run without entry.

Importantly, note that it is sufficient for the result in Proposition 3 if gross profits are only weakly decreasing with the number of firms. That is, even if $\pi(n+1) = \pi(n)$ for $n > 1$, the strict inequality between equilibrium investments holds if $n > \tilde{n}$. The intuition for this is a survival contest, the drivers of which are independent of current profits. To see this, consider a simple homogeneous-product Bertrand competition, where $\pi(n+1) = \pi(n) = 0$ for all $n > 1$. In this case, the investment in quality (moral hazard) is driven by future expectations on market structure, namely by the possibility that the firm may become a monopolist. However, the further away the current market structure from the monopoly state (the bigger n), the less likely the firm will be the lucky one to become a monopolist since too many firms compete for the "prize". Therefore the expected value of the continuation payoff is smaller, leading to less investment. As the market becomes concentrated, the stakes rise since the probability that the firm will be the one which survives into the monopoly stage increases. As a consequence, firms care more about future and invest more. Thus, as the market is driven towards concentration, the survival contest becomes more intense, firms grow more desperate to survive and as a result will fail with smaller probability. In general a firm in a more concentrated market will invest more, even if current gross profits are no more than that of less concentrated markets ($\pi(n+1) = \pi(n)$) resulting paradoxically in firms in more concentrated markets being less profitable.

Competition has two conflicting effects on the investment in quality. On the one hand, through price competition, it depresses profits and undermines the incentive to invest. On the other hand, it allows consumers to switch and hence to inflict maximum punishment on firms. Maximum punishment in and of itself, however, fails to give firms sufficient incentives to invest in quality if the market is too fragmented. Interestingly, as market concentration increases, the consequences of maximum punishment become more severe even if prices do not increase with concentration. This leads to firms investing more and, consequently, less moral hazard.

In the banking industry, for instance, numerous studies have found a negative relationship between concentration and deposit rates (see Brewer III and Jackson III (2006) and references therein). Carrying on interpreting the price game as a competition for deposits, this observation is equivalent to a decreasing $\pi(n)$. However, Brewer III and Jackson III (2006) find empirical evidence that although banks in concentrated markets pay lower deposit rates, they also take less risk. In other words, deposit rates are lower in more concentrated markets and these depressed deposit rates allow banks to invest in less risky assets (i.e. assets with smaller returns). In the present framework, this "sacrifice of potential profits" is $g(x)$, the opportunity cost of investing in less risky assets. Proposition 3, therefore, confirms the rationale behind the recent empirical findings of Brewer III and Jackson III (2006). In sum, while market power may increase gross profits, these profits are (partly) used to finance investment in quality (i.e. less risky assets).

The simple structure of the model allows me to fully characterise the equilibria by comparative static analysis. The following Proposition summarizes these results.

Proposition 4 *(Comparative statics) In a symmetric equilibrium without entry, the higher the per period gross profit $\pi(n)$, the monopoly profit $\pi(1)$, or the discount factor β, the higher amount firms invest in quality, the lower the moral hazard. That is,*

$$\frac{dx(n)}{d\pi(n)} > 0, \frac{dx(n)}{d\pi(1)} > 0, \frac{dx(n)}{d\beta} > 0$$

Proof. See Appendix B. ■

The effects of the parameters are very intuitive. Higher current profits increase the value of being in the game so firms care more about survival and invest more in quality. In addition, more patient firms are more concerned with the future and this also leads to more investment. Surprisingly, higher monopoly profit increases the amount of investment and, consequently, reduces the moral hazard in *all* market structures. As a result, it is possible to discipline oligopolies with the mere potential that they can become a monopoly. That is, the higher the value of monopoly, the more oligopolists will invest in quality, reducing the chance that the market actually ever becomes a monopoly. Regardless of how low the value of gross profit $\pi(n)$ is, as the value of monopoly tends to infinity, the oligopolists' investment will approach the maximum investment in quality and the market turns into a monopoly with probability approaching zero. In other words, as the monopoly profit tends to infinity ($\pi(1) \rightarrow \infty$), any market structure can be sustained almost surely, regardless of the intensity of price competition. That is, firms engage in fierce competition to survive and as a consequence, none will fail with probability approaching one, preserving current market structure almost surely. Note that if the monopoly profit is sufficiently large, then firms are willing to endure severe losses in the oligopoly phase, even if there is little hope that the monopoly stage will ever be reached. The following Corollary which is a direct consequence of Proposition 4 summarises this observation.

Corollary 1 *In a symmetric equilibrium of the game without entry, as the value of the monopoly firm approaches infinity, the optimal investment in quality tends to the maximum value. That is,* $\lim_{\pi(1)\rightarrow\infty} x(n) = 1$*. As a consequence, for any* n*, the expected number of failures tends to zero:* $\lim_{\pi(1)\rightarrow\infty} n(1 - x(n)) = 0$.

The survival contest interpretation of the game, the "contest for a prize" has some features in common with patent races. In both models, the higher the prize, the more firms invest. However, there is at least one crucial difference between the current model and a standard patent race. The survival contest identified in the current model is not a race, in fact it's quite the opposite. In a simple symmetric patent race more investment in the probability of discovery would imply a quicker end of the game since the probability that one of the firms wins is increasing with investments. In the current model, however, higher level of investment increases the probability of firm survivals and as such *reduces* the probability that any of the firms succeeds in winning the prize (e.g. becomes a monopolist in the Bertrand case). In other words, the probability that the "game ends" and the prize will be won is decreasing with the value of the prize, which is quite the opposite to what we observe in patent race models. This implies that the survival contest is a much more ruthless struggle since firms have to endure investment costs for a much longer period.

2.2 The game with entry

In this section, I explore industry dynamics in a more realistic setting incorporating the possibility of entry. I introduce entry in the simplest possible way modelling it as free and deterministic. The main insights of the previous section readily carry through so I omit rigorous proofs for the sake of brevity.

There is an infinite number of potential entrants who have access to the same technology and they are completely aware of the industry dynamics. The fixed cost of entry, $F > 0$, is commonly known. There is free and non-stochastic entry: a potential entrant will enter the market as long as the expected value of the firm is higher than the cost of entry. Assumption 1-3. are maintained throughout this section. Timing within a period is modified as follows:

1. Entry decisions and (sequential) entry.

2. Given the number of firms present in the market (i.e. incumbents plus entrants) n, profits, $\pi(n)$, are realised.

3. Firms simultaneously choose investment in quality, x.

4. Failures and exits occur.

Consider the following simple dynamic programme of potential entrants:

$$\widehat{V}^e(n) = \pi(n) + \max_{0 \leq x \leq 1} -g(x) + \beta x \widehat{V}^e(n) \tag{4}$$

where the e superscript stands for entry. $\widehat{V}^e(n)$ are equilibrium firm values for which entrants entertain the idea of entry. To see the rationale behind this programme, consider the subgame when the number of incumbents is n', and a potential entrant is contemplating entry. Note the stationarity of the problem: if it is profitable to enter today for a market structure $n' + 1$ (incumbents plus the entrant), then it will be profitable for other entrants to enter tomorrow for market structures $n \leq n' + 1$. Hence, the continuation value in (4) is not a function of market structures $n < n' + 1$. Define $\bar{n}$ such that

$$\bar{n} = \begin{cases} 0 \text{ if } \widehat{V}^e(1) < F \\ n : \widehat{V}^e(n) \geq F > \widehat{V}^e(n+1) \text{ otherwise} \end{cases} \tag{5}$$

By Assumption 2, it is immediate from the envelope theorem that $\widehat{V}^e(\cdot)$ is decreasing everywhere, therefore $\bar{n}$ is uniquely defined. It also follows from (5) that entry will never challenge incumbents if and only if $F > \widehat{V}^e(2)$, regardless of the value of monopoly.

First consider the subgame, when the market structure is $\bar{n}$. At this point, no entrant will enter because the newcomer in the market with $\bar{n} + 1$ firms (incumbents plus the entrant) will make a loss. However, if firms failed at the end of the last period so that the current period starts with $n < \bar{n}$ incumbents, then (sequential) entry immediately follows until the number of firms (i.e. incumbents plus entrants) is $\bar{n}$ again. Note that, therefore, (4) is not a function of rivals' investments any more in contrast to the dynamic programme (1). When the number of firms is $\bar{n}$, it doesn't matter how many firms will fail at the end of the period since at the beginning of the next period entry immediately fills in the gap brought along by failures: this is the result of free entry. This effectively creates an upper bound on concentration. That is, if $x(\bar{n}) < 1$, then firms will fail with positive probability but in equilibrium exactly as many firms enter as fail leaving the market structure unchanged.[11] Therefore, in the subgame with $n \leq \bar{n}$ firms in the market, the following strategies constitute an equilibrium:

Equilibrium Strategy of Potential Entrants: if the number of firms is $n < \bar{n}$, then enter, otherwise stay out.

Equilibrium Strategy of Incumbents and Entrants: choose x so as to maximise the programme (4).

Now consider the subgame $n > \bar{n}$. For this, the dynamic programme is as follows:

$$v^e(n; x_{-i}) = \pi(n) + \max_{0 \leq x_i \leq 1} \left\{ -g(x_i) + \beta x_i \left[\sum_{k=0}^{n-\bar{n}-1} V^e(n-k) \Pr(k|x_{-i}) + \widehat{V}^e(\bar{n}) \sum_{k=n-\bar{n}}^{n-1} \Pr(k|x_{-i}) \right] \right\} \quad (6)$$

where again $\Pr(k|x_{-i})$ is the probability mass function of the convolution of $n-1$ Bernoulli trials and $V^e(n-k)$ is the equilibrium value of the firm when there are $n-k$ firms in the market. Now the continuation payoff consists of two parts. The second sum accounts for the expected value of the states $n \leq \bar{n}$. As discussed above, in this case free entry ensures that $n = \bar{n}$ at the beginning of each period, that is, if the previous period ended with $n - k \leq \bar{n}$ firms in the market, the next period price game will start with $\bar{n}$ firms. The first sum accounts for the expected value of continuation payoffs for market structures $n - k > \bar{n}$, where entry is not profitable, hence does not occur. The programme is straightforward to solve recursively as before. Let the solution to the programme (6) in a symmetric equilibrium be $x^e(n)$. The next Proposition is the analogue of Proposition 3.

Proposition 5 *In a symmetric equilibrium of the game with entry, more firms in the market implies strictly less investment in quality, that is,* $x^e(n+1) < x^e(n)$*, unless* $n+1 < \bar{n}$*, in which case* $x^e(n+1) = x^e(n) = 1$.

Proof. See Appendix B. ■

[11] Note that this essentially results in a degenerate steady state distribution. However, the distribution of failures (or entries) is *not* degenerate in this steady state.

Observe that if $\bar{n}$ is such that the equilibrium investment in quality is $x^e(n) = 1$ for some $n \geq \bar{n}$ (that is, $\tilde{n} \geq \bar{n}$), then the equilibrium market structure with entry is $n_E^* \in [\bar{n}, \tilde{n}]$, at which no firm fails and no firm enters. In other words, in this case all states $\bar{n} \leq n \leq \tilde{n}$ are absorbing. However, if $x^e(n) < 1$ for all $n \geq \bar{n} > 0$ (that is, $\tilde{n} \leq \bar{n}$), then the market is in the state of continuous turbulence in the steady state, $n_E^* = \bar{n}$. In this case, exactly as many firms enter as fail in each period and the expected number of failures (entrants) in the steady state is equal to $\bar{n}(1 - x^e(\bar{n}))$. This can be readily summarised in the following Proposition.

Proposition 6 *In a symmetric equilibrium of the game with entry, the long run market structure is $n_E^* \in [\bar{n}, \tilde{n}]$ if $\tilde{n} > \bar{n}$ and $n^* = \bar{n}$ if $\tilde{n} \leq \bar{n}$. The expected number of failures in the steady state is equal to $n_E^*(1 - x^e(n_E^*))$.*

Note that if the nature of price competition is homogeneous-product Bertrand, that is $\pi(n) = 0$ for all $n > 1$, then entry is profitable if and only if there is no firm in the market, regardless of how large the monopoly profit and entry cost are. In this case, once competition is introduced, the market will steadily converge towards monopoly and remains a monopoly forever. However, the higher the monopoly profit, the slower the convergence is as it is highlighted in Corollary 1. Perhaps paradoxically, cut-throat rivalry, therefore, leads to monopoly even with free entry, regardless of entry costs and how big the monopoly profit is.

It is easy to see that entry makes everything worse in terms of expected quality. Entrants make it impossible for the market to reach the most valuable states, therefore they forcibly reduce the expected continuation payoff and, as a result, the value of the firm at all market structures. As a result, entry further depresses the level of investment in quality since firms care less about the future.

Corollary 2 *In a symmetric equilibrium, the value of the firm and investment in quality is smaller with entry than without entry. That is, $V^e(n) < V(n)$ and $x^e(n) < x(n)$ if $\bar{n} > 1$.*

Proof. It is easy to see that the programmes (1) and (6) are identical if $\bar{n} = 1$. From this and Assumption 2 it follows that $\widehat{V}^e(\bar{n}) < V(1)$ for $\bar{n} > 1$. Then a straightforward inductional argument implies that $\widehat{V}^e(\bar{n}) < V(n)$ for $\bar{n} \geq n$, which in turn implies $V^e(n) < V(n)$ for all n. But then, as in Proposition 3, the FOCs imply $x^e(n) < x(n)$. ■

3 Discussion of results

In this paper, market structure dynamics plays the key role in the determination of investment in quality, which, in turn, fuels the evolution of market dynamics. Competition and unforgiving consumers can depress prices to the level where investment is insufficient to prevent failure ($x < 1$). This mechanism triggers failures setting the market in motion. Market dynamics are driven by firms' expectations of future market structure, regardless of price dynamics. Firms want to make it into some profitable market structure but the further they are from this state, the less probable they will ever reach it. As a result, they have little to gain from survival, consequently care less about the future and invest too little. The closer the "winners" get to a profitable stage, however, the more desperate they become to survive because their stake in survival is much higher: it is more likely they will be among the few, who actually arrive at the lucrative phase. In other words, some win, others lose and failing firms indirectly raise the stakes for the survivors. Surviving firms, therefore, end up in a fierce contest to stay on and invest more in quality. Paradoxically, this reduces the chance that the desired stage will ever be reached. As a result, fragmented markets display heavy shake-out and fewer and fewer failures occur as the market gets concentrated.

Intense competition evolves into market concentration. Circumstances that make competition tougher have the effect of blocking the natural evolution of market dynamics. Entry, for instance, has the potential to prevent high quality and exacerbate moral hazard even in the long run. With entry, incumbents may never reach sufficiently profitable stages, therefore, they always underinvest keeping the market in ongoing turbulence.

The main result of this study is therefore quite different from previous findings in the literature of reputation and competition: for example in Allen (1984) and Hörner (2002), high quality is sustainable in competitive markets when quality is imperfectly observed. The main driving force behind their result is the sophisticated consumer whose behaviour consists of two fundamental building blocks. First, as in my model, switching consumers rationally inflict maximum punishment on firms in equilibrium, should they ever experience bad quality. Second, crucially, sophisticated consumers know the technology in addition and hence they can pin down firms through an incentive compatibility constraint that leads consumers to deducing unobserved quality from observable prices. Thus the maximum punishment strategy of unforgiving consumers has supreme bite in their models due to the signalling role of prices. I construct an equilibrium in which price does not have this signalling role and consumers form their expectations based only on firms' past quality performance. Note that this simple consumer behaviour is consistent with unsophisticated consumers who don't know the technology but it is also consistent with sophisticated consumers in a pooling equilibrium when firms' prices convey no information to consumers. I show that consumers' maximum punishment strategy (Assumption 1) in and of itself fails to give the right incentives for firms to produce high quality in general. Market evolution, however, will lead to high quality equilibrium in the long run if it exists. Observe that the very nature of strategic gambling for survival (i.e. firms optimally choose $x < 1$) ensures steady convergence towards high quality. It is probably instructive to replicate Allen's results in the present framework. Should consumers be completely aware of industry dynamics, they could calculate the threshold in (3). If the value of high quality is sufficiently high (and/or utility from bad quality is sufficiently low), then consumers would be willing to pay a price at any given market structure which is just high enough to get firms' gross profit above this threshold. In this case, we end up in a stationary equilibrium, just like in Allen's study: each firm invests the maximum amount ($x = 1$), no firm fails and price is above marginal cost. Lastly, my model shows some similarity to that of Rob and Fishman (2005). In their model, firms are local monopolists and therefore there is no strategic interaction among them. Reputation about firms spreads among consumers by word of mouth, securing a continuous customer base expansion for surviving firms. This expanding demand boosts period monopoly profits, which results in an investment pattern increasing with firm age. Therefore in their article, similarly to my result, high quality is a product of some sort of evolution.

4 Conclusion

In this paper, I have developed a model in which strategic rivalry leads firms to underinvest in imperfectly observed quality (moral hazard). Consumers follow a strategy by which they inflict maximum punishment abandoning the firm should it deliver low quality. This is the standard reputational mechanism through which asymmetric information problems are supposed to get rectified. In my model, however, this very mechanism in and of itself is not sufficient in general to give firms the necessary incentives to produce high quality. Rivalry triggers a survival contest and failing firms drive the market towards a concentrated structure. I show that the moral hazard problem is effectively eased along this dynamics even when price does not increase with concentration. In particular, since market concentration increases the prospects for firms to reach profitable states, the evolution of industry dynamics induces higher investments in quality reducing moral hazard. However, perhaps paradoxically because higher concentration leads to a more intense contest to survive, it also reduces the probability that these much wanted future profitable states will ever be reached.

The model explains why, even in the absence of scale economies, we may observe massive failure rates and steady convergence towards concentration in fragmented markets of some experience and credence goods. It also explains why more concentrated markets produce higher qualities in general, regardless of the intensity of price competition. The study, therefore, hopes to contribute to our understanding of the peculiar market transformation and dynamics in industries of such a particular importance as banking, health care and audit. A notable policy implication is that in some markets concentration seems to be necessary to alleviate moral hazard. As a consequence, promoting competition in markets where imperfectly observed product characteristics are of utmost importance to society may well be counterproductive.

Appendix A The game with consumers: an extension

In this appendix, I discuss a straightforward way to include consumers into the game analysed in the body of the paper. The main objective is to show that consumers' maximum punishment strategy is rational and can emerge as an equilibrium outcome of an extended game, similarly to Allen (1984), Hörner (2002) and Rob and Fishman (2005). Assumption 1 and 2 are satisfied in the equilibrium of the game.

As in Salop (1979), there is a continuum of consumers, whose measure is normalised to one, located uniformly along a circle. Each consumer buys one product each period. The product can be of two qualities, high or low. Quality is not observed before purchase but after consumption. The value of this unobserved quality to each consumer is identical, $u > 0$ if quality is high and zero if quality is low. Products are horizontally differentiated. Denote the distance between a consumer and firm i by $s_i \in (0, 1)$.

The *ex post* net benefit for a buyer to consume the good from firm i at time t and price $p_i^t \in \mathbb{R}_+$ is $U_i^t = u - p_i^t - cs_i^t$ if quality is high, and $U_i^t = -p_i^t - cs_i^t$ if quality is low. The parameter $c \geq 0$ measures the travel cost, and is an index of the degree of product differentiation. If $c = 0$, then goods are homogeneous, whereas if $c > 0$ products are differentiated. This utility function implies identical consumer attitudes towards (unobserved) quality, since every consumer gets negative utility from consuming a low quality product at a positive price, regardless of her distance from firms. Were the quality known before purchase, no consumer would buy low quality products and high quality producers would face a downward sloping demand schedule. The *ex ante* net benefit to consume the good produced by firm i at time t is $EU_i^t = q_i^t u - p_i^t - cs_i^t$, where q_i^t is the consumer's belief of buying a high quality product from firm i.

Suppose in addition that u is sufficiently high so that consumers always buy if $q_i^t > 0$.

At time t there are n^t firms symmetrically located on the circle.[12] Let $N^t = \{1, ..., n^t\}$ be the set of firms in period t. The cost of production is zero. In each period, firm i can invest x_i^t in quality at a cost of $g(x_i^t)$. After investing x_i^t firm i produces low quality with probability $1 - x_i^t$. The per period profit net of investment for firm i is $\hat{\pi}(p_i^t; p_{-i}^t, \mu^t)$, where $p_{-i}^t = [p_j^t]_{j \neq i}$ and μ^t is a vector of consumers' actions defined below.

Timing

In each period, decisions and actions take place in the following order:

1. Firms decide to quit or stay. Incumbents relocate symmetrically.
2. Firms choose prices simultaneously.

[12] It is possible to extend the model to allow firms to choose their location. Economides (1984) shows that with free location choice, there exists a symmetric equilibrium in locations and prices.

3. Consumers choose firms on the basis of last period quality, travel distance and prices.
4. Firms choose investment in quality simultaneously.
5. Consumers and firms observe current period quality.

Markov Strategies

I will look at Markov strategies that only depend on last period qualities. Define the (last period) history of qualities as: $H^t = \times_{i \in N^{t-1}} \{u^{t-1}, 0\}$. A consumer chooses which firm to buy from, as a function of the history of qualities, her distance from firms and current prices. Her strategy is described by the mapping $\mu : H^t \times [\times_{i \in N^t}(0,1)] \times [\times_{i \in N^t}\mathbb{R}_+] \rightarrow N^t$. Having observed last period qualities, firms decide to quit or stay, then incumbents choose prices. After having observed its own gross profits from the price game, a firm sets the level of investment in quality. Thus, a firm i's strategy consists of three mappings: $\tau_i : H^t \rightarrow \{Quit, Stay\}$, $p_i : H^t \rightarrow \mathbb{R}_+$ and $x_i : H^t \times \mathbb{R}_+ \rightarrow [0,1]$.

Symmetric Markov Perfect Equilibrium

The following strategies and beliefs constitute a Symmetric Subgame Perfect Nash Equilibrium in Markov strategies:

Equilibrium Strategy of Consumers:

- Do not buy from a firm, which has produced bad quality.[13]
- Buy from a firm which minimizes total costs, $p_i^t + cs_i^t$.

Equilibrium Beliefs of Consumers:

- If a firm has produced bad quality, it will always produce bad quality with probability one.
- At period t, firms produce good quality with probability $q_i^t = q^t = x_*^t$, if they haven't produced bad quality before.

Equilibrium Strategy of Firms:

- Quit if you produced bad quality last period.
- If you produced good quality last period, choose p^t such that $p^t = r(p_{-i}^t, \mu^t) \in \arg\max_{p_i^t} \hat{\pi}(p_i^t; p_{-i}^t, \mu^t)$.

[13] This is Assumption 1 in the main text.

- If you produced good quality last period, choose x^t which maximises the dynamic programme of (1), where $\pi(n) \equiv \hat{\pi}(r(p_*^t, \mu_*^t), p_*^t, \mu_*^t)$.

It is simple to see why p_*^t, x_*^t are optimal for firm i, and also why firm i is no better off staying in the market if it has produced bad quality last period, given consumers' and other firms' strategy. Looking at (1) and Proposition 4, observe that there is no dynamic link between price and investment in quality: given other firms' and consumers' strategy (i.e. that they do not trade off price and quality), posting a price other than p_*^t would just reduce gross profit, which would in turn induce lower investment in quality, lowering the overall value of the firm.

On the other hand, given firms' strategy, consumers can clearly do no better by switching to another firm because, by the symmetric nature of the equilibrium, consumers would be no better off if $c = 0$, and would be worse off if $c > 0$. Also, consumer can do no better by deviating to staying loyal rather than switching from the firm which has produced bad quality. Finally, note that a Nash equilibrium in Markov strategies is necessarily subgame perfect.

Appendix B Proofs

The following Lemmas will be useful in what follows.

Lemma B.1 *$xg'(x) > g(x)$ for $x > 0$.*

Proof. Recall that $g''(\cdot) > 0$ by Assumption 3. Hence, the following holds for any two distinct points v, x in the domain: $g(v) > g(x) + g'(x)(v - x)$. Letting $v = 0$ and recalling that $g(0) = 0$ gives the inequality result. ∎

Lemma B.2 *In a symmetric equilibrium, the value function is strictly decreasing: $V(n) < V(n-1)$.*

Proof. The proof is by induction. First, I show that $V(2) < V(1)$. Then supposing $V(n) < ... < V(1)$, I'll show that $V(n+1) < V(n)$ which proves the claim. To prove $V(2) < V(1)$ observe that

$$\begin{aligned}
V(2) - V(1) &= \pi(2) - g(x(2)) + \beta x(2)[V(2)x(2) + V(1)(1 - x(2))] - V(1) \\
&= \pi(2) - g(x(2)) + \beta x(2)V(1) + \beta x(2)^2[V(2) - V(1)] - V(1) \\
&= \frac{\pi(2) - g(x(2)) + \beta x(2)V(1) - V(1)}{1 - \beta x(2)^2} < \frac{\pi(1) - g(x(2)) + \beta x(2)V(1) - V(1)}{1 - \beta x(2)^2} \leq 0
\end{aligned}$$

where the first inequality follows from Assumption 2 and the second from the fact that $\pi(1)-g(x(2))+\beta x(2)V(1) \leq V(1)$ since $x(2)$ is the maximiser of the duopoly's dynamic programme, rather than the monopoly's.

Next, suppose $V(n) < \ldots < V(1)$. Then, I'll show that $V(n+1) < V(n)$. Let $x \equiv x(n+1)$ be the value which maximises the firm's dynamic program in a symmetric equilibrium when there are $n+1$ firms in the market. Then,

$$V(n+1) = \pi(n+1) - g(x) + \beta \sum_{k=0}^{n} V(n+1-k)\binom{n}{k}(1-x)^k x^{n+1-k}$$

Therefore,

$$\begin{aligned} V(n+1) \quad \leq \quad & \pi(n) - g(x) + \beta \sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^k x^{n-k} \\ & + \beta \sum_{k=0}^{n} V(n+1-k)\binom{n}{k}(1-x)^k x^{n+1-k} \\ & - \beta \sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^k x^{n-k} \end{aligned} \tag{B.1}$$

where the inequality follows from Assumption 2 again. Now, observe that

$$\pi(n) - g(x) + \beta \sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^k x^{n-k} \leq V(n)$$

since x maximises the value function when there are $n+1$ firms, rather than n. Thus, from (B.1) we get

$$\begin{aligned} V(n+1) - V(n) \quad \leq \quad & \beta \sum_{i=-1}^{n-1} V(n-i)\binom{n}{i+1}(1-x)^{i+1} x^{n-i} \\ & - \beta \sum_{i=0}^{n-1} V(n-i)\binom{n-1}{i}(1-x)^i x^{n-i} \end{aligned}$$

where I switched indexes setting $k = i+1$ in the first term and $k = i$ in the second. I use the convention $\binom{n}{l} = 0$ for $l < 0$ and $l > n$ throughout the derivations. Next, using Pascal's identity

$$\binom{n}{i+1} - \binom{n-1}{i} = \binom{n-1}{i+1} \tag{B.2}$$

the inequality can be rearranged as

$$\begin{aligned}
V(n+1)-V(n) &\leq \beta\sum_{i=-1}^{n-1} V(n-i)\left[\binom{n-1}{i+1}+\binom{n-1}{i}\right](1-x)^{i+1}x^{n-i} \\
&\quad -\beta\sum_{i=0}^{n-1} V(n-i)\binom{n-1}{i}(1-x)^{i}x^{n-i} \\
&= \beta\sum_{i=-1}^{n-1} V(n-i)\binom{n-1}{i+1}(1-x)^{i+1}x^{n-i} \\
&\quad -\beta\sum_{i=0}^{n-1} V(n-i)\binom{n-1}{i}(1-x)^{i}x^{n+1-i}
\end{aligned}$$

The last element in the first sum after equality can be dropped since it's zero. Then, switching the index back again, set $i=k-1$ in the first sum and $i=k$ in the second. This yields:

$$\begin{aligned}
V(n+1)-V(n) &\leq \beta\sum_{k=0}^{n-1} V(n+1-k)\binom{n-1}{k}(1-x)^{k}x^{n+1-k} \\
&\quad -\beta\sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^{k}x^{n+1-k} \\
&= \beta\sum_{k=0}^{n-1}[V(n+1-k)-V(n-k)]\binom{n-1}{k}(1-x)^{k}x^{n+1-k}
\end{aligned}$$

Rearranging yields

$$V(n+1)-V(n) \leq \frac{\beta\sum_{k=1}^{n-1}[V(n+1-k)-V(n-k)]\binom{n-1}{k}(1-x)^{k}x^{n+1-k}}{1-\beta x^{n+1}} < 0$$

where the last inequality holds by the induction hypothesis. ■

Corollary B.1 *Suppose $x(n)<1$. Then, in a symmetric equilibrium, the value of the firm when there are n firms in the market is $V(n)=\pi(n)-g(x(n))+x(n)g'(x(n))>\pi(n)$. Furthermore, $0<V(n)<\infty$.*

Proof. The first part of the claim is the result of the first order condition (2) being substituted into the value function; the inequality follows from Lemma B.1. The second part of the Corollary, $0<V(n)$ holds because $0\leq\pi(n)$ by Assumption 2 and the strictly increasing function $-g(x)+xg'(x)>0$ for all $x>0$ by Lemma B.1. $V(n)<\infty$ since $V(n)<V(1)$ for all $n>1$ by Lemma B.2 and $V(1)\leq\pi(1)/(1-\beta)<\infty$.

■

PROOF OF PROPOSITION 1

By inspection of (1), it is immediate that Blackwell's sufficiency conditions, namely monotonicity and discounting, hold. Therefore, the value function is unique. Also, the second order condition is simply

$-g''(x) < 0$, which holds by Assumption 3. This implies that the reaction functions are unique. It remains to show that these reaction functions define a symmetric equilibrium of the game in pure strategies. This follows from the the symmetric reaction functions being continuous and downward sloping. Let's write the first order condition to the programme (1) in the following way

$$f_i(x_i; x_j, x_l, n) \equiv \beta x_j \sum_{k=0}^{n-2} V(n-k)c_k + \beta(1-x_j)\sum_{k=0}^{n-2} V(n-1-k)c_k - g'(x_i) = 0$$

where $x_l = [x_h]_{h \neq i,j}$ and $c_k = \Pr(k|x_l)$ is the probability mass function of the convolution of $n-2$ Bernoulli trials so $\sum_{k=0}^{n-2} c_k = 1$. Continuity is obvious. Furthermore, the reaction functions are downward sloping because, by the Implicit Function Theorem,

$$\frac{\partial x_i}{\partial x_j} = -\frac{\partial f_i(x_i; x_j, x_l, n)}{\partial x_j} / \frac{\partial f_i(x_i; x_j, x_l, n)}{\partial x_i} = \frac{\beta \sum_{k=0}^{n-2}[V(n-k) - V(n-1-k)]c_k}{g''(x_i)} < 0$$

where the inequality follows from Lemma B.2. Note that the reactions functions are linear ($\partial x_i / \partial^2 x_j = 0$), hence the symmetric equilibrium is unique if $|\partial x_i / \partial x_j| \neq 1$.

PROOF OF PROPOSITION 3

The proof consists of two parts. In the first part I show that the left hand side of the first order condition decreases with x. In the second, I prove that it also decreases stochastically with n. As a result, higher n must imply lower x in equilibrium.

Let's define the left hand side of the FOC (2) as follows:

$$f(x; n, \beta) \equiv \beta \sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^k x^{n-k-1} - g'(x) \tag{B.3}$$

First, I show that $\partial f(x; n, \beta)/\partial x < 0$. To see this, differentiate (B.3)

$$\frac{\partial f(x; n, \beta)}{\partial x} = \beta \frac{\partial}{\partial x} \sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^k x^{n-1-k} - g''(x)$$

The last term is negative by Assumption 3. The first term is proportional to

$$\begin{aligned} \frac{\partial}{\partial x} \sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^k x^{n-1-k} &= \sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(n-1-k)(1-x)^k x^{n-2-k} \\ &\quad - \sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k} k(1-x)^{k-1} x^{n-1-k} \end{aligned}$$

Dropping the $(n-1)$th element in the first sum and the first element in the second sum, which are both zero, we obtain

$$\begin{aligned}\frac{\partial}{\partial x}\sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^k x^{n-1-k} &= (n-1)\sum_{k=0}^{n-2} V(n-k)\frac{(n-2)!}{k!(n-2-k)!}(1-x)^k x^{n-2-k} \\ &\quad -(n-1)\sum_{k=1}^{n-1} V(n-k)\frac{(n-2)!}{(k-1)!(n-1-k)!}(1-x)^{k-1}x^{n-1-k}\end{aligned}$$

Substitute $i=k$ in the first sum and $i=k-1$ in the second sum. Then

$$\begin{aligned}\frac{\partial}{\partial x}\sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^k x^{n-1-k} &= (n-1)\sum_{i=0}^{n-2} V(n-i)\frac{(n-2)!}{i!(n-2-i)!}(1-x)^i x^{n-2-i} \\ &\quad -(n-1)\sum_{i=0}^{n-2} V(n-1-i)\frac{(n-2)!}{i!(n-2-i)!}(1-x)^i x^{n-2-i} \\ &= (n-1)\sum_{i=0}^{n-2}[V(n-i)-V(n-1-i)]\binom{n-2}{i}(1-x)^i x^{n-2-i} \\ &< 0\end{aligned}$$

because from Lemma B.2 $V(n-i) < V(n-1-i)$ for all $i = 0,..n-2$. Therefore, we've established that

$$\frac{\partial f(x;n,\beta)}{\partial x} < 0 \tag{B.4}$$

Next, I'll show that $f(\cdot)$ is (stochastically) decreasing in n. In particular, for any $x \in (0,1)$, $f(x;n+1,\beta) < f(x;n,\beta)$. For this, it is sufficient to show that

$$\begin{aligned}\sum_{k=0}^{n} V(n+1-k)\binom{n}{k}(1-x)^k x^{n-k} &< \sum_{i=0}^{n-1} V(n-i)\binom{n-1}{i}(1-x)^i x^{n-1-i} \\ &= \sum_{k=1}^{n} V(n+1-k)\binom{n-1}{k-1}(1-x)^{k-1}x^{n-k}\end{aligned} \tag{B.5}$$

where the equality follows from setting $i=k-1$. In other words, the expected value of $V(n+1) < \ldots < V(1)$ with CDF $B(v;n,1-x)$ is smaller than the expected value of $V(n) < \ldots < V(1)$ with CDF $B(v;n-1,1-x)$, where:

$$
\begin{aligned}
B(v;n,1-x) &= \sum_{k=0}^{v} b(k;n,1-x) = \sum_{k=0}^{v} \binom{n}{k} (1-x)^k x^{n-k} \\
B(v;n-1,1-x) &= \sum_{k=1}^{v} b(k;n-1,1-x) = \sum_{k=1}^{v} \binom{n-1}{k-1} (1-x)^{k-1} x^{n-k}
\end{aligned}
$$

Now, since $V(\cdot)$ is strictly increasing in k (Lemma B.2), by the First Order Stochastic Dominance Theorem (FOSD), (B.5) holds if and only if $B(v;n-1,1-x)$ with support $v=1,...,n$ (first order) stochastically dominates $B(v;n,1-x)$ with support $v=0,...,n$; that is,

$$
B(v;n,1-x) \geq B(v;n-1,1-x) \tag{B.6}
$$

for $v=0,...,n$ and with strict inequality for some v. Inequality (B.6) is shown to hold by induction. It can be easily seen that

$$
\begin{aligned}
B(0;n,1-x) - B(0;n-1,1-x) &= x^n \\
B(1;n,1-x) - B(1;n-1,1-x) &= (n-1)(1-x)x^{n-1}
\end{aligned}
$$

Therefore, the inductional hypothesis is

$$
B(v;n,1-x) - B(v;n-1,1-x) = \binom{n-1}{v} (1-x)^v x^{n-v} \tag{B.7}
$$

Observe that

$$
B(v+1;n,1-x) = B(v;n,1-x) + b(v+1;n,1-x)
$$

Therefore,

$$
\begin{aligned}
B(v+1;n,1-x) - B(v+1;n-1,1-x) &= B(v;n,1-x) - B(v;n-1,1-x) \\
&\quad + b(v+1;n,1-x) - b(v+1;n-1,1-x)
\end{aligned}
$$

Thus, using the inductional hypothesis (B.7), equation above can be written as

$$
\begin{aligned}
B(v+1;n,1-x) - B(v+1;n-1,1-x) &= \binom{n-1}{v}(1-x)^v x^{n-v} \\
&\quad + \binom{n}{v+1}(1-x)^{v+1}x^{n-v-1} - \binom{n-1}{v}(1-x)^v x^{n-v-1} \\
&= \binom{n-1}{v}(1-x)^v x^{n-v-1}[x-1] + \binom{n}{v+1}(1-x)^{v+1}x^{n-v-1} \\
&= \left[\binom{n}{v+1} - \binom{n-1}{v}\right](1-x)^{v+1}x^{n-v-1} \\
&= \binom{n-1}{v+1}(1-x)^{v+1}x^{n-v-1}
\end{aligned}
$$

where in the last line I used Pascal's identity (B.2). Therefore, the inductional hypothesis is proven, and we can conclude that for all v

$$
B(v;n,1-x) - B(v;n-1,1-x) = \binom{n-1}{v}(1-x)^v x^{n-v} \geq 0
$$

The inequality is strict for all $v < n$. Therefore, by the theorem of FOSD, inequality (B.5) holds, and we can conclude that since $f(x;n,\beta)$ is decreasing in both x and n. The first order condition (B.3), therefore, implies that in a symmetric equilibrium the more firms are in the market, the less they invest; in other words, $x(n+1) < x(n)$ as required.

PROOF OF PROPOSITION 4

Recall the first order condition (B.3)

$$
f(x;n,\beta) = \beta \sum_{k=0}^{n-1} V(n-k)\binom{n-1}{k}(1-x)^k x^{n-k-1} - g'(x) = 0
$$

Then,

$$
\begin{aligned}
\frac{\partial f(x;n,\beta)}{\partial \pi(n)} &= \beta x^{n-1}\frac{\partial V(n)}{\partial \pi(n)} = \frac{\beta x^{n-1}}{1-\beta x^n} > 0, \\
\frac{\partial f(x;n,\beta)}{\partial \pi(1)} &= \beta \sum_{k=0}^{n-1} \frac{\partial V(n-k)}{\partial \pi(1)}\binom{n-1}{k}(1-x)^k x^{n-k-1} > 0 \ .
\end{aligned}
$$

The second inequality follows from a simple induction argument. Clearly,

$$
\frac{\partial V(1)}{\partial \pi(1)} = \frac{1}{1-\beta x_1} > 0
$$

Now suppose $\partial V(1)/\partial \pi(1), ..., \partial V(n-1)/\partial \pi(1) > 0$. Then,

$$\frac{\partial V(n)}{\partial \pi(1)} = \frac{\beta}{1-\beta x^n} \sum_{k=1}^{n-1} \frac{\partial V(n-k)}{\partial \pi(1)} \binom{n-1}{k} (1-x)^k x^{n-k} > 0$$

where the inequality holds by the induction hypothesis. So

$$\sum_{k=0}^{n-1} \frac{\partial V(n-k)}{\partial \pi(1)} \binom{n-1}{k} (1-x)^k x^{n-k} > 0$$

A very similar argument shows that

$$\frac{\partial f(x;n,\beta)}{\partial \beta} > 0,$$

Then, using the Implicit Function Theorem, (B.4) and the inequalities derived above, we have

$$\begin{aligned} \frac{dx(n)}{d\pi(n)} &= -\frac{\partial f(x;n,\beta)}{\partial \pi(n)} / \frac{\partial f(x;n,\beta)}{\partial x} > 0, \quad \frac{dx(n)}{d\pi(1)} = -\frac{\partial f(x;n,\beta)}{\partial \pi(1)} / \frac{\partial f(x;n,\beta)}{\partial x} > 0, \\ \frac{dx(n)}{d\beta} &= -\frac{\partial f(x;n,\beta)}{\partial \beta} / \frac{\partial f(x;n,\beta)}{\partial x} > 0 \end{aligned}$$

as required.

PROOF OF PROPOSITION 5

The case of the corner solution with $n+1 < \bar{n}$ is obvious. The rest of the proof is identical to that of Proposition 3 after proving the monotonicity of the value function. The proof of the value function's strict monotonicity is very similar to that of Lemma B.2, therefore I only show the first step of the induction argument applied there. Rewriting (6) in a symmetric equilibrium for $\bar{n}+1$ and letting $x \equiv x^e(\bar{n}+1)$

$$\begin{aligned} V^e(\bar{n}+1) - \widehat{V}^e(\bar{n}) &= \pi(\bar{n}+1) - g(x) + \beta x V^e(\bar{n}+1)\Pr(0|x) + \beta x \widehat{V}^e(\bar{n}) \sum_{k=1}^{\bar{n}} \Pr(k|x) - \widehat{V}^e(\bar{n}) \\ &= \pi(\bar{n}+1) - g(x) + \beta x V^e(\bar{n}+1) x^{\bar{n}} + \beta x \widehat{V}^e(\bar{n})(1-x^{\bar{n}}) - \widehat{V}^e(\bar{n}) \\ &= \pi(\bar{n}+1) - g(x) + \beta x \widehat{V}^e(\bar{n}+1) - \beta x[\widehat{V}^e(\bar{n}+1) - \widehat{V}^e(\bar{n})] \\ &\quad + \beta x^{\bar{n}+1}[V^e(\bar{n}+1) - \widehat{V}^e(\bar{n})] - \widehat{V}^e(\bar{n}) \\ &\leq \widehat{V}^e(\bar{n}+1) - \widehat{V}^e(\bar{n}) - \beta x[\widehat{V}^e(\bar{n}+1) - \widehat{V}^e(\bar{n})] + \beta x^{\bar{n}+1}[V^e(\bar{n}+1) - \widehat{V}^e(\bar{n})] \\ &= \frac{1-\beta x}{1-\beta x^{\bar{n}+1}} [\widehat{V}^e(\bar{n}+1) - \widehat{V}^e(\bar{n})] < 0 \end{aligned}$$

The first inequality follows from the fact that x is a maximiser of $v^e(\cdot)$, rather than $\widehat{v}^e(\cdot)$. The last inequality follows from the definition of $\bar{n}$ since it implies $\widehat{V}^e(\bar{n}) > \widehat{V}^e(\bar{n}+1)$. Then, substitute $\widehat{V}^e(\bar{n})$

for $V(n-k)$ for $n-k \leq \bar{n}$ and $V^e(n-k)$ for $V(n-k)$ for $n-k > \bar{n}$ in the derivations of Lemma B.2 and the result is immediate. But then the strict monotonicity of $x^e(n)$ readily follows from an argument identical to the one in the proof of Proposition 3.

Chapter III

Market Structure Dynamics with Moral Hazard and Stochastic Entry

The stage game is defined among n incumbents and N (*ex ante* heterogeneous) potential entrants in an infinitely repeated framework and the dynamics of market structure is investigated. The equilibrium is a Markov chain where transition probabilities are functions of incumbents' and entrants' strategies yielding a completely endogenous stochastic structure. Incumbents can invest in their survival while facing competition within and outside the market. Regardless of the intensity of competition, moral hazard drives the market towards concentration, while concentration effectively eases the moral hazard problem. However, the increasing threat of entry upsets future prospects, dampening the effect of the survival contest at work. As a result, concentration can never fully resolve the problem of moral hazard resulting in a steady state market structure of ongoing turbulence. The model exhibits entry and exit, shake-outs from time to time, and correlated exit and entry rates in the limit.

1 Introduction

In this paper, I introduce stochastic entry in the model studied in Chapter II. The presence of stochastic entry serves two purposes. First, stochastic entry is a natural robustness check; it permits me to show that the main results of the model with deterministic entry are not fragile. Second, stochastic entry allows the study of more complex steady state dynamics.

In Chapter II, the presence of moral hazard coupled with strategic interaction among firms results in firms gambling on their survival. Since quality is unobservable before consumption, firms have an incentive to cheat on consumers (moral hazard). In equilibrium, firms invest too little in high quality provision when market concentration is low. Consumers, on the other hand, abandon firms which produce low quality, setting the market in motion. As failing firms drive the market towards concentration, survivors invest more along the equilibrium path, even when prices do not increase with concentration. This results in a monotonic convergence towards concentration, although at an ever decreasing speed. Entry, of course, can potentially block this monotonic market evolution. In Chapter II, I introduced entry into the game in the simplest possible way, modelling it as free and deterministic. As a consequence, the stochastic structure was particularly simple, due to the one way stochastic process. The deterministic nature of entry defined an entry point above which entry was profitable and therefore free entry would effectively block any further concentration. This coupled with the monotonic convergence towards concentration naturally led to a degenerate steady state distribution of the market structure. In each period, as many firms enter as they have failed, resulting in a market structure at a standstill in the long-run equilibrium. Therefore, although the distribution of the number of failures (entry) in the steady state is not degenerate, the steady state distribution of the number of firms in the market is degenerate in that model.

It is natural to extend the results and introduce stochastic entry in the model. The reason for this extension is essentially threefold. First, it is important to investigate whether the strong monotonicity of the equilibrium investment profile is sensitive to the assumption of an extremely simple entry process. Thus, one may conjecture that the results of Chapter II are not robust, since the monotonicity of the continuation value may break down depending on the distribution of entry. In this Chapter, I demonstrate that this is not the case; even with stochastic entry, the equilibrium investment schedule remains strongly monotone. Second, as discussed above, a rather unappealing feature of deterministic entry is that it results in a degenerate steady state. In this Chapter, I show that with stochastic entry there exists a non-degenerate steady state distribution, which yields a more complex and interesting steady state dynamics. Finally, in this Chapter, the stochastic entry process is endogenous; that is, the probabilities of entry are determined in equilibrium. The game therefore yields a Markov chain with a completely endogenous stochastic structure. This is a realistic feature of the model which differentiates it from most of the literature on industry dynamics. Even though many of the properties of the steady state distribution of market structure cannot be explored analytically, important insights can be gained. The model displays shake-outs from time to time and correlated exit and entry rates. These two empirical regularities are often rationalised using technological arguments, whereas in my model they are simply the results of moral hazard.

1.1 Previous literature

Surprisingly, there are only a handful of studies which fully investigate market structure dynamics with stochastic entry. Most game theoretic studies only address stochastic industry dynamics in a duopoly-monopoly framework. The few papers that go beyond this simplification and assume an arbitrary number of firms often abstract from some important dimensions of the strategic interaction among firms. For instance, it is often the case that incumbents do not act strategically, or that the stochastic nature of entry is exogenous. Finally, articles which analyse industry dynamics in the framework of a fully fledged stochastic game are almost never tractable analytically. I will now briefly discuss the seminal articles of the literature on stochastic entry. My review is not meant to be exhaustive; I will focus on studies which have an arbitrary (finite or infinite) number of firms.

Deshmukh and Chikte (1976) develop a simple model where incumbents do not act strategically, there is no exit, and entry is governed by a Poisson process. The intensity parameter of the Poisson process is a function of the current market price. In their model incumbents simply share the industry profit symmetrically, and hence have a natural incentive to increase price. A high price increases the intensity parameter of the Poisson process; that is, it attracts more entry. The lack of strategic interaction among incumbents leads to surprising results. When only few firms share the industry profit, the marginal effect of entry on the incumbent's profit is relatively big and hence firms have more of an incentive to deter entry and decrease price. In contrast, when the market is shared by many, one additional firm would not lower an individual firm's profit much and so entry deterrence is less important. As a result, the number of firms steadily converges to infinity, while the price tends to the monopoly price. Despite the lack of strategic product market interaction, in this model entry is truly stochastic in the sense that it is a random process governed by market fundamentals and the entrant has a direct effect on these fundamentals in equilibrium.

Jovanovic (1982) introduces a model of industry dynamics with heterogeneous firms. He too abstracts from strategic interaction among firms; he casts his analysis in a perfectly competitive environment. Due to aggregate certainty, the price follows a deterministic path in equilibrium. The costs on the other hand are random. There is no technological progress, costs are a function of a random shock and of the efficiency of the firm ("type"). The firm, however, does not know its efficiency level but infers it over time through the sequence of realised costs. In other words, some firms are (stochastically) more efficient than others and they learn their efficiency rank ("type") over time. Based on the stochastic technology and the deterministic price sequence, firms make decisions on entry, exit and production. The equilibrium leads to selection: efficient (low cost) firms grow and survive while inefficient ones disappear. As a result, the average profit typically increases as the industry matures. In the limit, the model exhibits no entry and exit, a somewhat unappealing feature.

Hopenhayn (1992) focuses on stationary equilibrium in a model of a competitive industry with productivity shocks and heterogeneous firms. Similarly to Jovanovic (1982), prices follow a deterministic path in equilibrium, so the only source of uncertainty is productivity. Hopenhayn simplifies Jovanovic's structure by discarding the idea of efficiency types. Hence, there is no firm level learning and selection process which would lead to a degenerate steady state distribution as in Jovanovic (1982). Instead, Hopenhayn introduces idiosyncratic productivity shocks that evolve as a Markov process. As a consequence, his model yields a stationary long-run equilibrium with entry and exit, and hence a proper steady state distribution of market structure that he fully characterizes with comparative static analysis.

In Jovanovic (1982) and Hopenhayn (1992), stochastic entry doesn't really add much to the stochastic structure of the models, due to the absence of strategic interaction among firms. Entrants do not learn their productivity shocks before entering and, as a result, in both models entry is a deterministic function of the state in equilibrium. In addition to prices (which follow a deterministic - hence completely foreseen - path in both models), the state is defined as the firm's belief over its efficiency level in Jovanovic (1982), and the firm's current productivity shock in Hopenhayn (1992). In equilibrium, therefore, entry does not affect the state of the individual decision maker. In sum, given the state of the industry, entry is deterministic and doesn't affect incumbents' decisions, due to the perfectly competitive environment.

Ericson and Pakes (1995) are the first to address market structure dynamics with heterogeneous firms and stochastic entry in a full (i.e. non duopoly) game theoretic framework. A firm's payoff and its value are a function of its own and its rivals' efficiency levels. Hence the state of the game is the vector of efficiency levels. The efficiency levels, in turn, depend on the incremental investments as well as idiosyncratic and aggregate random shocks. Idiosyncratic shocks lead to gradual decay in firm efficiency in the absence of investment, whereas aggregate shocks induce correlation in profits (an empirical regularity). In every period, firms decide to enter, stay on/exit and how much to invest, the latter being private information. The results can be summarized as follows. There is a closed set of states which is absorbing: the market evolves into this set of states over time and never leaves it. Within this set, typically all sorts of individual firm and industry dynamics are possible in the steady state. Furthermore, the investment schedule is usually not monotonic in the state. Indeed, if a firm is highly efficient it optimally chooses not to invest, since the marginal benefit is outweighed by the marginal cost of investment. Similarly, at low efficiency levels firms do not invest, again because their instantaneous profits are too low. In this model, entry is truly stochastic; although the number of entrants is a deterministic function of the state in equilibrium, entry does affect the state of the game, namely the distribution of firm efficiencies.

Amir and Lambson (2003) develop a model of a market with strategic interaction but simplify the rich stochastic structure of Ericson and Pakes (1995) in many ways. The stochastic structure of the game is completely exogenous; the market conditions evolve according to an exogenous (although not necessarily Markovian) random process. To these market conditions, firms react by taking actions which are simply whether to be active in a period. Firms play a symmetric game and exit is governed by a last-in-first-out rule. In this model, similarly to Jovanovic (1982) and Hopenhayn (1992), entry is only stochastic because the state of the game evolves stochastically. In equilibrium, entry is a deterministic function of the state of the game, and hence affects the state of the game (the number of active firms) deterministically.

In this Chapter, I develop a model where entry is completely stochastic. In particular, the number of entrants in a given period is a stochastic function of the state of the game and therefore entry affects the state stochastically. The stochastic entry process is the result of *ex ante* heterogenous entrants. Before entry, entrants observe their own fixed entry cost which is random and private information. This generates a non-degenerate distribution of the number of entrants at each state, as opposed to a degenerate one as in previous studies. As a result, the entry process stochastically affects the state of the game which is the number of firms or market structure.

2 The model

Time is discrete and the common discount factor is $\beta \in (0,1)$. The infinitely repeated stage game consists of two types of rounds: in each period there are N entry followed by one production rounds. That is, in every period there are $N+1$ rounds. In each entry round one entrant is allowed to enter, in other words there are N potential entrants who enter sequentially. Time "stops" within the period, there is discounting only across periods.

Let the fixed cost of entry be $F = \bar{F} + \varepsilon$, where $\bar{F} > 0$ and ε is iid with $E(\varepsilon) = 0$ and CDF $\rho(\cdot)$ with support $[-\bar{F}, \infty]$. Each entrant knows her own fixed cost of entry *before* entering. An entrant's fixed cost is private information, but all entrants know the distribution of fixed costs. In other words, entrants are heterogenous. This property complicates the space of strategic interactions since, in addition to the game between incumbents and entrants, it allows for a non-trivial game among the entrants themselves. In particular, entrants are not symmetric, so despite the sequential structure, they are unable to foresee the entry process with certainty within a period. As a result, upon entry an entrant must contemplate the fact that further entry in that period can make her entry unprofitable *ex post*. In previous studies, the symmetry of entrants (before entering) and the sequential structure often led to entry being a deterministic function of the state of the game in equilibrium. In this model, on the contrary, the n firms in the market in round l (that is, the incumbents and the entrants who have entered before round $l+1$) will expect an entrant to enter in the next round with probability $\rho_{l+1,n}$, which is to be determined endogenously. In sum, heterogeneity before entry induces a non-degenerate distribution of the number of entrants at each state.

Assumption 1. *In each period there are N potential entrants, which enter sequentially at the beginning of the period. The l-th entrant enters with probability $\rho_{l,n}$ when there are n firms in the market at the end of the previous (production or entry) round.*

After the N entry rounds, the production round starts. At the beginning of each production round, firms engage in price (or quantity) competition. I denote the symmetric equilibrium profit for a firm from the price game by $\pi(n)$, when n firms are in the market.

Assumption 2. $\pi(n) \leq \pi(n-1)$, *for* $n > 2$ *and* $\pi(2) < \pi(1)$, *where* $\pi(\cdot) \geq 0$. *Also,* $\lim_{n \to \infty} \pi(n) = 0$.

Note that Assumption 2 is general enough to account for virtually all forms of imperfect competitions usually considered in the literature.

Next, having realised the gross profits from the price game, firms invest in quality the amount of which is private information (moral hazard). The probability of producing high quality in a given production round depends on how much firms invest in that round. In particular, firm i can invest in quality $x_i \in [0, 1]$ at a cost of $g(x_i)$ in a production round and produces high quality at the end of that round with probability x_i.

Assumption 3. *If a firm produces low quality, it fails.*

Assumption 3 reflects that consumers stop patronising a firm if a firm has cheated on them. In the literature on reputation, this is a common consumer behaviour in equilibrium. It is often considered to be the very minimum requirement for consumer rationality and most models demand more sophisticated consumers. However, I exclude consumers from the current game for the sake of simplicity, hence the (minimal) assumption on their behaviour. As a consequence of Assumption 3, a firm fails with probability $1 - x_i$ at the end of the production round. Therefore, including firm i, the number of firms (the market structure) at the end of the production round will be $n-k$ with probability $x_i \Pr(k|x_{-i})$, where $x_{-i} = [x_j]_{j\neq i}$ and $k = 0, ..., n-1$ is the number of firms other than firm i failing at the end of the production round. $\Pr(k|x_{-i})$ is the probability mass function of the convolution of $n-1$ Bernoulli distributions with "success" probabilities x_{-i}.

Therefore, the net profit is $\pi(n) - g(x_i)$ in a period. I make the following assumptions on the cost function:

Assumption 4. $g(x_i), g'(x_i), g''(x_i) > 0$ *for* $x_i > 0$ *and* $g(0) = g'(0) = 0$, $\lim_{x\to 1} g(x) = \infty$.

The assumptions on the cost function are very standard: the costs are strictly increasing and convex. The limit on the cost function is just to rule out corner solutions for the sake of simplicity.

The summary of timing within a period is as follows:

1. The first entrant observes the number of firms in the market after failures occurred in the last period and enters with probability $\rho_{1,n-k}$. Subsequently, the next entrants observe the number of firms in the last entry round and enter sequentially. The lth entrant, therefore, enters with probability $\rho_{l,n}$ if there are n firms in the $l-1$th entry round.

2. After the last entry round, the production round starts. Given the number n of active firms (entrants plus incumbents), gross profits, $\pi(n)$, are realised.

3. Firms simultaneously choose investment in quality, x.

4. Failures, exits occur.

The strategies are assumed to be Markov and I focus on symmetric Markov Perfect Equilibrium (MPE). Let the equilibrium value function at the end of the lth entry round be $W_l(n)$. (That is, n includes the lth entrant if it has entered the market) Then,

$$W_l(n) = W_{l+1}(n+1)\rho_{l+1,n} + W_{l+1}(n)(1-\rho_{l+1,n}) \tag{1}$$

where $\rho_{l+1,n}$ is the probability that in the next entry round the $l+1$th entrant will enter when there are n firms in the market. Note that at the end of the last entry round, since there is no further entry during that period, $W_N(n) = V(n)$ where $V(n)$ is the equilibrium value function of the firms at the beginning of the production round. The incumbents' value function in symmetric equilibrium then is defined by

$$V(n) = \pi(n) - g(x_n) + \beta x_n \sum_{k=0}^{n-1} [W_1(n-k+1)\rho_{1,n-k} + W_1(n-k)(1-\rho_{1,n-k})] \binom{n-1}{k} (1-x_n)^k x_n^{n-1-k} \tag{2}$$

where x_n is the optimal level of investment, i.e. the solution to the dynamic program (2) when there are n firms in the market.

Proposition 1 *$V(n)$ and $W_l(n)$ are strictly decreasing in n.*

Proof. The proof is as follows. First, I consider the finite version of the dynamic programme and by a straightforward inductional argument I establish that $V_t(n-1) - V_t(n) \geq 0$ and $W_{l,t}(n-1) - W_{l,t}(n) \geq 0$ for all t, l and n, where $V_t(\cdot)$ and $W_{l,t}(\cdot)$ are value functions when there are t periods remaining. Then, by standard dynamic programming argument $\lim_{t\to\infty} V_t(n) = V(n)$ so it is proven that $V(n-1) - V(n) \geq 0$. Similarly $W_l(n-1) - W_l(n) \geq 0$. Lastly, I will show that these inequalities must be strict.

In what follows I drop the index of x_n in order to ease notation so $x \equiv x_n$. By Assumption 2, it is trivial to see that $V_0(n-1) - V_0(n) \geq 0$ for all n since $V_0(n) = \pi(n)$. Now noting that $V_0(n) = W_{N,0}(n)$, from (1) a straightforward backwards induction shows that $W_{l,0}(n-1) - W_{l,0}(n) \geq 0$ for all l and n. So suppose $V_{t-1}(n-1) - V_{t-1}(n) \geq 0$ and $W_{l,t-1}(n-1) - W_{l,t-1}(n) \geq 0$ for all l and n. Then, from (2) and Assumption 2 again

$$\begin{aligned} V_t(n) \leq\ & \pi(n-1) - g(x) \\ & + \beta x \sum_{k=0}^{n-1} [W_{1,t-1}(n-k+1)\rho_{1,n-k} + W_{1,t-1}(n-k)(1-\rho_{1,n-k})] \binom{n-1}{k} (1-x)^k x^{n-1-k} \end{aligned} \quad (3)$$

Now observe that if x is the maximiser of $V_t(n)$, then

$$\begin{aligned} & \pi(n-1) - g(x) \\ & + \beta x \sum_{k=0}^{n-2} [W_{1,t-1}(n-k)\rho_{1,n-k-1} + W_{1,t-1}(n-k-1)(1-\rho_{1,n-k-1})] \binom{n-2}{k} (1-x)^k x^{n-2-k} \\ \leq\ & V_t(n-1) \end{aligned}$$

Therefore, (3) can be rearranged easily as

$$[V_t(n) - V_t(n-1)]/\beta x \le$$

$$\begin{aligned}
&\sum_{k=0}^{n-1}[W_{1,t-1}(n-k+1)\rho_{1,n-k} + W_{1,t-1}(n-k)(1-\rho_{1,n-k})]\binom{n-1}{k}(1-x)^k x^{n-1-k} \\
&-\sum_{k=0}^{n-2}[W_{1,t-1}(n-k)\rho_{1,n-k-1} + W_{1,t-1}(n-k-1)(1-\rho_{1,n-k-1})]\binom{n-2}{k}(1-x)^k x^{n-2-k} \\
= &\sum_{i=-1}^{n-2}[W_{1,t-1}(n-i)\rho_{1,n-i-1} + W_{1,t-1}(n-i-1)(1-\rho_{1,n-i-1})]\binom{n-1}{i+1}(1-x)^{i+1} x^{n-2-i} \\
&-\sum_{i=0}^{n-2}[W_{1,t-1}(n-i)\rho_{1,n-i-1} + W_{1,t-1}(n-i-1)(1-\rho_{1,n-i-1})]\binom{n-2}{i}(1-x)^i x^{n-2-i}
\end{aligned}$$

where after the last equality sign in the first sum I substituted $k = i+1$ and in the second $k = i$. Using Pascal's identity

$$\binom{n-1}{i+1} - \binom{n-2}{i} = \binom{n-2}{i+1}$$

and noting that by convention $\binom{n}{i} = 0$ for $i < 0$ and $i > n$, the inequality can be rearranged again as

$$[V_t(n) - V_t(n-1)]/\beta x \le$$

$$\begin{aligned}
&\sum_{i=-1}^{n-2}[W_{1,t-1}(n-i)\rho_{1,n-i-1} + W_{1,t-1}(n-i-1)(1-\rho_{1,n-i-1})][\binom{n-2}{i+1} + \binom{n-2}{i}](1-x)^{i+1} x^{n-2-i} \\
&-\sum_{i=0}^{n-2}[W_{1,t-1}(n-i)\rho_{1,n-i-1} + W_{1,t-1}(n-i-1)(1-\rho_{1,n-i-1})]\binom{n-2}{i}(1-x)^i x^{n-2-i} \\
= &\sum_{i=-1}^{n-2}[W_{1,t-1}(n-i)\rho_{1,n-i-1} + W_{1,t-1}(n-i-1)(1-\rho_{1,n-i-1})]\binom{n-2}{i+1}(1-x)^{i+1} x^{n-2-i} \\
&-\sum_{i=0}^{n-2}[W_{1,t-1}(n-i)\rho_{1,n-i-1} + W_{1,t-1}(n-i-1)(1-\rho_{1,n-i-1})]\binom{n-2}{i}(1-x)^i x^{n-1-i}
\end{aligned}$$

Switching the indices back again, this leads us to

$$\begin{aligned}
V_t(n) - V_t(n-1) &\le \beta\sum_{k=0}^{n-2}([W_{1,t-1}(n-k+1)\rho_{1,n-k} + W_{1,t-1}(n-k)(1-\rho_{1,n-k})] \\
&\quad -[W_{1,t-1}(n-k)\rho_{1,n-k-1} + W_{1,t-1}(n-k-1)(1-\rho_{1,n-k-1})])\binom{n-2}{k}(1-x)^k x^{n-k} \\
&\le 0 \qquad (4)
\end{aligned}$$

where the last inequality follows from the induction hypothesis: since $W_{1,t-1}(n-k+1) \le W_{1,t-1}(n-k) \le W_{1,t-1}(n-k-1)$ for all $0 \le k \le n-2$, all the elements of the sum on the right hand side of (4) are non-positive. Noting that $V_t(n) = W_{N,t}(n)$, from (1) a straightforward application of backwards induction shows that $W_{l,t}(n-1) - W_{l,t}(n) \ge 0$. Therefore, for all t, l and n, $V_t(n-1) - V_t(n) \ge 0$ and

$W_{l,t}(n-1) - W_{l,t}(n) \geq 0$. Taking the limit $\lim_{t\to\infty} V_t(n) = V(n)$, it follows that $V(n-1) - V(n) \geq 0$ and similarly, $W_l(n-1) - W_l(n) \geq 0$ for all n and l. Weak monotonicity, therefore is established. In what follows, I show that these two weak inequalities are actually strict and hence the strong monotonicity of $V(n)$ and $W_l(n)$.

Observe that inequality (4) holds at the limit too, that is, we can consider it without the time subscripts. Furthermore, from Assumption 2 for $n = 2$ inequality (3) is actually strict, therefore, (4) can be written as

$$V(2) - V(1) < \beta x^2[W_1(3)\rho_{1,2} + W_1(2)(1-\rho_{1,2})] - [W_1(2)\rho_{1,1} + W_1(1)(1-\rho_{1,1})] \leq 0$$

The second inequality follows from what we have established earlier, that is $W_l(n-1) - W_l(n) \geq 0$ for all l. But then $W_{N-1}(1) = V(2)\rho_{N,1} + V(1)(1-\rho_{N,1}) > W_{N-1}(2) = V(3)\rho_{N,2} + V(2)(1-\rho_{N,2})$ since $W_N(n) = V(n)$. Noting again $W_l(2) - W_l(3) \geq 0$ for all l, this implies through backwards induction that $W_1(1) > W_1(2)$. As a consequence, there is at least one negative element on the right hand side of (4), which just means the second inequality in (4) is strict. Therefore, $V(n) < V(n-1)$ for all n. Following the logic of the argument above it immediately follows that $W_l(n) < W_l(n-1)$ for all l and n. ∎

Having established the strong monotonicity of the value function now I can prove the strong monotonicity of the equilibrium investments.

Proposition 2 *Firms' investment level is a strictly increasing function of market concentration, that is* $x_{n+1} < x_n$.

Proof. I'll show that the first order condition decreases in both x and n (stochastically). This implies that whenever n increases x must decrease. Consider the first order condition of the dynamic programme (2) in symmetric equilibrium:

$$f(x;n,\beta) \equiv \beta \sum_{k=0}^{n-1} [W_1(n-k+1)\rho_{1,n-k} + W_1(n-k)(1-\rho_{1,n-k})] \binom{n-1}{k} (1-x)^k x^{n-1-k} - g'(x) = 0$$

with solution x_n. Differentiating the FOC we get

$$\begin{aligned}
\frac{\partial f(x;n,\beta)}{\partial x} &= \beta\sum_{k=0}^{n-1}[W_1(n-k+1)\rho_{1,n-k}+W_1(n-k)(1-\rho_{1,n-k})]\binom{n-1}{k}(n-1-k)(1-x)^k x^{n-2-k} \\
&\quad -\beta\sum_{k=0}^{n-1}[W_1(n-k+1)\rho_{1,n-k}+W_1(n-k)(1-\rho_{1,n-k})]\binom{n-1}{k}k(1-x)^{k-1}x^{n-1-k}-g''(x) \\
&= \beta(n-1)\sum_{i=0}^{n-2}[W_1(n-i+1)\rho_{1,n-i}+W_1(n-i)(1-\rho_{1,n-i})]\frac{(n-2)!}{i!(n-2-i)!}(1-x)^i x^{n-2-i} \\
&\quad -\beta(n-1)\sum_{i=0}^{n-2}[W_1(n-i)\rho_{1,n-i-1}+W_1(n-i-1)(1-\rho_{1,n-i-1})]\frac{(n-2)!}{i!(n-2-i)!}(1-x)^i x^{n-2-i} \\
&\quad -g''(x)
\end{aligned}$$

where after the last equality I substituted $k=i$ in the first and $k=i+1$ in the second sum. Thus,

$$\begin{aligned}
\frac{\partial f(x;n,\beta)}{\partial x} &= \beta(n-1)\sum_{i=0}^{n-2}([W_1(n-i+1)\rho_{1,n-i}+W_1(n-i)(1-\rho_{1,n-i})] \\
&\quad -[W_1(n-i)\rho_{1,n-i-1}+W_1(n-i-1)(1-\rho_{1,n-i-1})])\binom{n-2}{i}(1-x)^i x^{n-2-i}-g''(x) \\
&< 0
\end{aligned}$$

where the last inequality follows from Proposition 1. It is not difficult to see that $f(x;n,\beta)$ is decreasing in n too (Full proof as in Proposition 3 in Chapter II). This implies the claim. ∎

It is important to note that the results in Proposition 1 and 2 do not depend on the distribution of entry probabilities $\rho(\cdot)$. Therefore, the strong monotonicity of the value functions and the investment schedule is completely invariant to the stochastic structure of the entry process.

I am now in the position to fully characterise the entry process. The lth entrant enters if $W_l(n+1) > \bar{F}$, that is, it enters with probability $\rho_{l,n} = \rho(W_l(n+1) - \bar{F})$ when there are n firms in the market at the end of the previous (production or entry) round. It is not difficult to see that $0 < W_l(n) < \infty$, therefore $0 < \rho_{l,n} < 1$. Consequently, entry is a continuous (stochastic) threat at all states. Since $W_l(n) < W_{l+1}(n)$, from (1) and Proposition 1 it is also simple to see that $\rho_{l,n} < \rho_{l+1,n}$ and $\rho_{l,n} > \rho_{l,n+1}$. In other words, within a period, *ceteris paribus* entrants enter with greater probability if closer to the production round and when the market is more concentrated. Both results are very intuitive. An entrant closer to the production round enters with bigger probability because the uncertainty regarding the remaining potential entries of that period is smaller. Also, more concentrated markets naturally attract more entry. There seems to be two countervailing effects. However, from (1) and Proposition 1 it is immediate that $\rho_{l,n} > \rho_{l+1,n+1}$. Therefore, if entry occurs in a given entry round, it will unambiguously lower the probability of further entry. These observations can be summarised in the following Proposition.

Proposition 3 $\rho_{l,n} < \rho_{l+1,n}$, $\rho_{l,n} > \rho_{l,n+1}$, $\rho_{l,n} > \rho_{l+1,n+1}$, $0 < \rho_{l,n} < 1$, $\lim_{n\to\infty} \rho_{l,n} = 0$.

Proof. The inequalities follow from Proposition 1 and the claim on the limit from Proposition 1 and Assumption 2. ■

The model describes a very intuitive industry dynamics. When there are many firms in the market, the equilibrium value of the firm is low. Thus, firms will invest very little and hence will fail with considerable probability. At the same time, the low value of the incumbent firms will attract little entry. As a result, the market is driven towards concentration. However, as the market gets concentrated, firms invest more and fail with smaller probability. In addition, since the value of being in the market increases, entry will occur with a greater probability preventing the market to become too concentrated. As a result, concentration can never fully resolve the moral hazard problem, since while concentration eases the problem of moral hazard, entry prevents the market to become too concentrated. Consequently, the model exhibits continuous turbulence in the form of failures and entry even in the limit.

Now I turn to the analysis of the long run behaviour of the model.

3 Steady State dynamics

The idea behind the analysis of the steady state dynamics is to use the equilibrium properties of the discrete time game above to derive transition probabilities when the period of the game is infinitely small. In essence, I am going to transform the discrete time Markov transition kernel into a continuous time analogue and will prove that the game produces a birth and death process.

Let's introduce explicitly the length of a period dt. Note that the results in the previous section were obtained in the special case when $dt = 1$. Since I am going to relate the dynamics to birth and death processes, it seems convenient to redefine the probability of failure as $1 - x_n = \theta_n$. Therefore, in a period a firm will fail with probability $\theta_n dt$ and will survive with probability $1 - \theta_n dt$. Also, the lth entrant enters with probability $\rho_{l,n} dt$ when there are n firms in the market. There are two facts that I am going to use extensively in what follows:

$$(dt)^i = o(dt) \text{ for } i \geq 2 \text{ and} \tag{5}$$

$$(1 - adt)^k = \sum_{i=0}^{k} \binom{k}{i} (-adt)^i = 1 - kadt + o(dt) \tag{6}$$

where $o(dt)$ is a term vanishing (in probability) at the rate of at least second order. Formally, $\lim_{dt \to 0} o(dt)/dt = 0$. Denote the random variable of the number of failures $\omega(t)$. Using (5) and (6) we can show that the transition probabilities in the production round are

$$\Pr(\omega(t+dt) = k|n) = \binom{n}{k}(\theta_n dt)^k (1-\theta_n dt)^{n-k} = \binom{n}{k}(\theta_n dt)^k (1-(n-k)\theta_n dt + o(dt))$$

$$= \begin{cases} 1 - n\theta_n dt + o(dt) & \text{if } k = 0 \\ n\theta_n dt + o(dt) & \text{if } k = 1 \\ o(dt) & \text{if } k > 1 \end{cases}$$

Similarly, letting the random variable $\psi(t)$ denote the number of entering firms, the transition probabilities in the entry process are

$$\Pr(\psi(t+dt) = j|n-k) = \begin{cases} \prod_{l=1}^{N}(1-\rho_{l,n-k}dt) = 1 - \sum_l \rho_{l,n-k}dt + o(dt) & \text{if } j = 0 \\ \sum_{l=1}^{N} \rho_{l,n-k}dt \prod_{i<l}(1-\rho_{i,n-k}dt) \prod_{i>l}(1-\rho_{i,n-k+1}dt) = \sum_l \rho_{l,n-k}dt + o(dt) & \text{if } j = 1 \\ o(dt) & \text{if } j > 1 \end{cases}$$

Let the random variable $Z(t)$ denote the number of firms in the market at time t. Then the change in $Z(t)$ is simply the convolution of $\omega(t)$ and $\psi(t)$. In particular, $dZ(t)/dt = \psi(t) - \omega(t)$. The transition probabilities in a period when there are n firms in the market are as follows

$$\begin{aligned}
\Pr(Z(t+dt) &= n+1|Z(t) = n) = [1 - n\theta_n dt + o(dt)][\textstyle\sum_l \rho_{l,n}dt + o(dt)] + o(dt) \\
&= \textstyle\sum_l \rho_{l,n}dt + o(dt) \\
\Pr(Z(t+dt) &= n-1|Z(t) = n) = [n\theta_n dt + o(dt)][1 - \textstyle\sum_l \rho_{l,n}dt + o(dt)] + o(dt) = n\theta_n dt + o(dt) \\
\Pr(Z(t+dt) &= n|Z(t) = n) = [1 - n\theta_n dt + o(dt)][1 - \textstyle\sum_l \rho_{l,n}dt + o(dt)] \qquad (7) \\
&\quad + [n\theta_n dt + o(dt)][\textstyle\sum_l \rho_{l,n}dt + o(dt)] + o(dt) \\
&= 1 - n\theta_n dt - \textstyle\sum_l \rho_{l,n}dt + o(dt) \\
\Pr(Z(t+dt) &= n+m|Z(t) = n) = o(dt) \quad \text{if } m \neq -1, 0, 1
\end{aligned}$$

with initial conditions $\Pr(Z(t+dt) = 0|Z(t) = 0) = 1 - \sum_l \rho_{l,0}dt + o(dt)$ and $\Pr(Z(t+dt) =$

$1|Z(t) = 0) = \sum_l \rho_{l,0} dt + o(dt)$. Denote the intensity parameters of the Markov process as $\mu_n = n\theta_n$ and $\lambda_n = \sum_l \rho_{l,n}$. These parameters are essentially the average number of failures ("death") and entry ("birth"), respectively. Although the entry process is sequential (a series of Bernoulli trials), it can be thought of as a binomial process with parameters N and $\overline{\rho_n}$ where $\overline{\rho_n} = \sum_l \rho_{l,n}/N$. Then it is easy to see that the process is essentially a birth-death process.

Proposition 4 *There exists a unique stationary steady state distribution P, with probability mass function $P_n = \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}P_0$ and $E(n) < \infty$.*

Proof. The process has a transition matrix

$$\mathbf{G} = \begin{pmatrix} 1-\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & 1-\lambda_1-\mu_1 & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & 1-\lambda_2-\mu_2 & \lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Stationarity implies that the average flow into a state is equal to the average flow out of that state. The stationary distribution, therefore, satisfies $P'\mathbf{G} = P$ where $P' = [P_0, P_1, ...]$. Thus

$$\begin{aligned} (1-\lambda_0)P_0 + \mu_1 P_1 &= P_0 \\ \lambda_{n-1}P_{n-1} + (1-\lambda_n-\mu_n)P_n + \mu_{n+1}P_{n+1} &= P_n \quad \text{for } n \geq 1 \end{aligned}$$

and solving this system of equations recursively yields

$$P_n = \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}P_0 \quad \text{for } n > 0$$

For existence the probability mass function must sum up to one, that is

$$\sum_{n=0}^{\infty} P_n = P_0 + \sum_{n=1}^{\infty} P_n = P_0 + P_0 \sum_{n=1}^{\infty}\prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}} = 1$$

Therefore, for the above to hold it is necessary and sufficient to show that

$$\sum_{n=1}^{\infty}\prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}} < \infty$$

From Proposition 3 $\lim_{n\to\infty}\lambda_n = 0$ and from Proposition 2 $\lim_{n\to\infty}\mu_n = \infty$. Applying, for instance, the ratio test it is immediate that the sequence converges (very rapidly, in fact). Then,

$$P_0 = \left(1 + \sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}\right)^{-1}$$

Uniqueness follows from the facts that the transition probability matrix is standard and the chain is irreducible. (Grimmett and Stirzaker 2001, Theorem 6.9.21) The chain is standard (although not uniform) because $\lim_{dt\to 0} \Pr(Z(t+dt) = i|Z(t) = j) = 1$ if $i = j$ and zero otherwise as it is clear from the definition of the transition probabilities in (7). The chain is clearly irreducible since any state can be visited from any state with strictly positive probability.

Furthermore,

$$E(n) = \sum_{n=1}^{\infty} nP_n < \infty$$

since applying for example the ratio test again

$$\lim_{n\to\infty} \frac{(n+1)P_{n+1}}{nP_n} = \lim_{n\to\infty} \frac{(n+1)\lambda_n}{n\mu_{n+1}} = \lim_{n\to\infty} \frac{\lambda_n}{n\theta_{n+1}} = 0$$

■

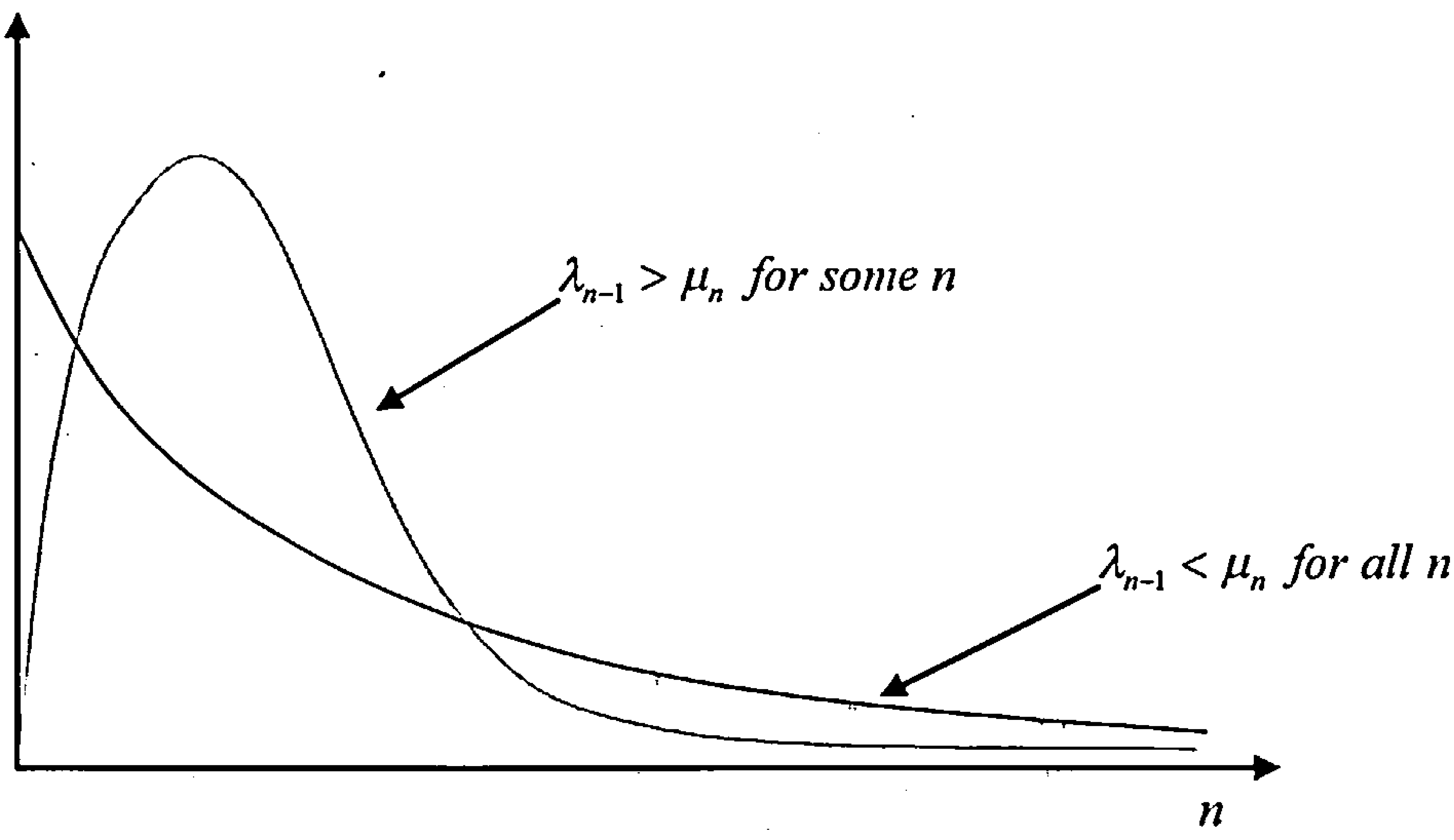
Proposition 5 *The most frequently visited state in steady state, that is, the mode of the steady state distribution is* $n^* = \arg\max\{P_n, P_{n-1} : \min_n |1 - \lambda_{n-1}/\mu_n|\}$.

Proof. This result is the direct consequence of the strong monotonicity of λ_n and μ_n. Observe that $P_n/P_{n-1} = \lambda_{n-1}/\mu_n$ and $\lim_{n\to\infty} \lambda_n = 0$ and $\lim_{n\to\infty} \mu_n = \infty$ so the probability mass function is increasing while $\lambda_{n-1} \geq \mu_n$ and decreasing afterwards. As a result, the probability mass function reaches it's maximum when the ratio λ_{n-1}/μ_n is closest to one. ■

4 Discussion

Unfortunately, it is not possible to derive closed form solutions for the statistics of the steady state distribution, such as the expected value or variance. Moreover, obtaining clear-cut comparative static analysis appears an impossible task. The reason is that parameters have an opposing effect on incumbents' and entrants' strategies. In particular, if for instance the instantaneous payoff $\pi(n)$ is higher for given n, then it has a positive direct effect on the continuation value in the obvious way, but it also increases the probability of entry, which lowers the continuation value. The sign is therefore ambiguous. Similarly for other parameters of the model.

Figure 1: Steady State Distributions



However, some properties of the distribution can be readily deduced. For instance, as a consequence of the strong monotonicity of the entry and failure parameters, the shape of the distribution can be of two types. If there exists no n such that $\lambda_{n-1} > \mu_n$, then the probability mass function is strictly decreasing and will look like an exponential distribution. In this case, the mode of the distribution, that is, the most frequently visited state is state zero (no firm in the market). However, if $\lambda_{n-1} > \mu_n$ for some n, then the probability mass function is strictly increasing for small n, and strictly decreasing afterwards. In other words, it will look like an F distribution. These two observations could be nicely described by the lognormal distribution: a high variance case will yield a shape similar to the exponential distribution, whereas a lower variance would result in a shape similar to the F distribution as depicted in Figure 1.

Using the recursive nature of the steady state probabilities $P_n/P_{n-1} = \lambda_{n-1}/\mu_n$ or $P_n - P_{n-1} = (\lambda_{n-1} - \mu_n)/\mu_n$, we can say something about the variance too. The distribution is tight (variance is small) if the ratio λ_{n-1}/μ_n is steep around n^*. In other words, if there is a big jump in the willingness to enter or in the failure rates around the most frequently visited state, then the variance will most probably be low. This may occur if the variance of the distribution of the fixed cost is relatively small or the cost function $g(\cdot)$ is relatively convex around n^*.

While the symmetric framework prohibits the model to address many interesting empirical phenomena (e.g. variance of firm growth rates, size distribution, etc.), it is still able to rationalise two important stylized facts, namely the presence of shake-out and the strong correlation between entry and exit rates.

Both are well known empirical regularities.

It has long been observed that most industries are characterised by shake-outs. In particular, in the period immediately after the birth of an industry there are only few firms. This is followed by a period of a steady increase in the number of incumbents. After a while, a wave of failures, mergers and acquisitions occurs. So far the theoretical literature explained shake-outs through the introduction of some innovation process (Sutton 2006). In the current model, a shake-out is simply the result of the presence of moral hazard, while the technology is constant over time. The model exhibits shake-outs from time to time: the market can end up with many firms with positive probability and then firms fail in a great number (high μ_n).

Entry and exit rates tend to be highly correlated in many industries. So far the literature has focused on the following possible driving forces: demand fluctuations, displacement of existing technologies, displacement of existing products, fluctuations in relative efficiency (Sutton 2006). The present model shows that correlation between entry and exit rates can also be just a direct consequence of moral hazard. This is immediate from Proposition 5 after observing that the most frequently visited state, the mode of the steady state distribution, is the market structure where entry and failure rates are approximately equal.

5 Conclusion

I have generalised the model in Chapter II by introducing stochastic entry. I have shown that the main result of that Chapter, the strong monotonicity of the equilibrium investment profile, is robust to the introduction of a stochastic entry process. In addition, I have proven the existence and uniqueness of a non-degenerate steady state distribution. In the most frequently visited state, the model exhibits continuous entry and failure in the limit. While it is not possible to fully characterise the steady state distribution with comparative static analysis in its general form, the model is capable to explain some important empirical regularities. In particular, moral hazard in and of itself is sufficient to produce shake-outs and correlated exit and entry rates.

Chapter IV

Testing Reputational Effects in a Natural Experiment: How did Arthur Andersen Clients Choose their New Auditor?

Reputation is the market mechanism which is widely believed to effectively resolve problems of asymmetric information. It relies, however, on consumers taking into account any available information on unobserved product characteristics (quality). Exploiting the natural experiment of Arthur Andersen's demise, I investigate to what extent if at all former Andersen clients took into consideration information on unobserved audit quality when they chose their new public accountant. In particular, the effect of auditors' financial restatement history on firms' auditor choice is examined in discrete choice settings. After controlling for auditor and client characteristics and for the endogeneity of audit fees, I find that although financial restatements are thought to be very noisy and often uninformative indicators of quality, former Andersen clients appear to have based their succeeding auditor decisions on them.

1 Introduction

In many markets, consumers are unable to ascertain before (and often after) consumption important product characteristics.[14] The firm, therefore, has a natural incentive to exploit its informational advantage and cheat on consumers. This is a classical problem of asymmetric information, a basic feature of numerous markets. A conventional solution to this problem is based on the recognition that market interactions are dynamic. In particular, if consumers suspect that they have been cheated, they will rationally abandon the firm: a dubious track record, i.e. "bad reputation" of the firm today can jeopardise its business tomorrow. Therefore, consumers' valuation of quality and informational asymmetry is at the heart of all reputational mechanism. This mechanism is effective as long as consumers are willing and able to read and correctly deduce information on quality from observables. In other words, reputation as a disciplinary force critically hinges on the assumption of consumer rationality.

Previous empirical studies have found mixed results regarding rationality of consumer behaviour calling into question the effectiveness of reputation as a disciplinary force in many markets. While it is clear that consumers value and act upon previously unavailable information on quality, it is less obvious if they are able and willing to decipher more complex quality signals. In other words, it appears unambiguous that consumers are not fools, but their strong rationality has not yet been decidedly confirmed in the empirical literature. This is a very important research area since in many markets only imperfect indicators are available for consumers to rely on in order to make reputation an effective solution to the widespread problem of asymmetric information.

In this paper, I analyse reputational effects in a market where customers are firms rather than individuals. Exploiting the natural experiment of Arthur Andersen's demise, I find that a noisy quality signal, the auditors' financial restatement history, is one of the most important driving forces in companies' auditor choice. This suggests that the reputational mechanism is indeed effective in the audit industry and clients are strongly rational.

In this study I investigate how former clients of Arthur Andersen chose their new auditor after Andersen's demise. In particular, I am interested in whether Andersen clients rationally took into account imperfect signals on quality when making their choice. The audit industry appears to be a particularly well suited laboratory to test buyer rationality for several reasons. First, moral hazard in this market is exceptionally severe: the existence of the whole industry hinges on the mechanism of reputation. Second, customers in this market are firms rather than individuals, therefore in general the strong notion of rationality should be found at work with a much better chance. Third, the reputation of the audit market as a whole had just suffered severe damage at the time of Andersen's failure and Andersen clients

[14] I will term these essential product characteristics as quality hereafter.

bore most of the consequences of the debacle. Therefore, they were probably much more sensitive to audit quality than other public companies: for them re-establishing the integrity of their financial statements by choosing auditor quality vigilantly was undoubtedly of utmost importance. Moreover, note that at the time simply choosing a big 4 auditor[15] could have appeared insufficient to signal integrity to investors since one of the big 5 had just failed. Perhaps not surprisingly there were widespread rumours at the time that Andersen's failure was not the only but merely the first one. (Economist, 2003) Therefore, it seems reasonable that former Andersen clients paid distinct attention to quality indicators that could potentially differentiate among auditors in general and within the exclusive group of big 4 in particular. Finally, Andersen's failure set the stage for a natural experiment. In many empirical works endogenous sample selection is a fundamental problem. That is, factors that induce selection into the sample may be correlated with variables which also drive the relationship of interest. In the current context, the analyst of auditor changes must be careful because (possibly unobserved) circumstances which trigger auditor switch may also affect the choice of the succeeding auditor. Focusing on a forced auditor change, however, sample selection that often confounds previous analyses is not an issue in the current empirical investigation.

The basic quality indicator in this study is auditors' financial restatement history. In particular, I investigate if former Andersen clients considered the proportion of an auditor's clients restating financial statements ("restatement history") as a quality signal while choosing among public accountants. Restatements are material corrections of financial reports which cannot be relied upon any more and, therefore, have to be reissued. Since the very role of the independent auditor is to review financial statements and ensure that they represent a fair account of the company's financial position, it appears natural to think that if a financial statement has to be restated, then the auditor has not done its job properly. A central question is, however, whether financial restatements can indeed be regarded as a genuine quality signal; that is, if they are correlated with audit quality at all. Since audit quality is never observed it is difficult to say anything explicit about this question. However, there are a number of reasons why it should be a valid quality indicator for companies. Firstly, and most importantly, capital markets do seem to perceive restatements as an important indicator of the authenticity of a company's financials. Financial statements are primarily produced and verified for shareholders. Consequently, public companies are concerned with the shareholders' perception of audit quality. And indeed, there is ample evidence that capital markets do receive restatements badly. Before Andersen's collapse, restatements hit the restating company hard: on average, they triggered a 10% fall in the company's market value between 1997 - 2002 (GAO-03-138). Therefore, companies should care about auditors' restatement profiles if only because

[15] The big 4 auditors are PricewaterhouseCoopers (PWC), Ernst&Young (EY), Deloitte&Touche (DT) and KPMG. These four auditors and Arthur Andersen constituted the big 5 before 31 August 2002.

capital markets do. Secondly, the Security and Exchange Commission (SEC) considers every restatement outright audit failure (GAO-03-138). Similarly, public bodies such as Financial Executives International (FEI) or Panel on Audit Effectiveness (O'Malley Panel) perceive restatements as an important indicator of audit quality (Palmrose and Scholz 2004). Consequently, there seem to be compelling reasons why financial restatement history should forcibly feed into companies' auditor choices. On the other hand, it is not quite clear to what extent the auditor is responsible for a restatement and as a consequence in the academic literature financial restatements are often thought to be very noisy (if at all) indicators of audit quality (Francis 2004). For instance, Eisenberg and Macey (2004) show that prior to Enron, there was no statistically significant difference across the big 5 auditors in terms of restatement history. In fact, Andersen outperformed, albeit marginally, some of the big 4 auditors on this measure. Also, between 1995 and 1999 only 13% of financial restatements led to auditor litigation. (Palmrose and Scholz 2004) Thus, it's not clear if and to what extent firms should take into account information on financial restatements when choosing an auditor. In this study I find that companies' auditor choices are robustly determined by auditors' financial restatement history and industry experience, whereas audit fees and company specifics have a very limited, often not detectable effect.

The research on the effectiveness of reputational mechanisms has very important policy implications. In the event of a major debacle, policy makers are often quick to declare market failure and step in by offering "effective regulations". For instance, Arthur Andersen's collapse was swiftly followed by the Sarbanes-Oxley Act which was supposed to rectify the audit market. However, in my view these debacles are more often than not part of a learning process and as such ultimately an efficient reputational mechanism. Indeed, these market crises draw the attention of agents at both sides of the market to the impact, the importance and ultimately the effectiveness of quality signals. Therefore, if a reputational mechanism is at work, then indirect measures based on strengthening existing market mechanisms (e.g. information disclosure) are probably much more beneficial than the costly and almost always handicapped direct regulation.

1.1 Previous empirical literature

Two strands of the empirical literature on unobserved quality can be distinguished. In one, consumer rationality is tested in a very obvious way: do consumers react and to what extent when previously unobserved quality becomes observable due to some information disclosure? In the other strand, a much stronger notion of rationality is analysed: are consumers able to and do they read potentially noisy and complex quality indicators?

The fundamental prediction has been readily confirmed on information disclosure: if consumers care about quality and quality gets revealed, then they should and indeed do take this information into consideration. Jin and Leslie (2003) find that once restaurant hygiene became easily observable as a result of a regulatory change, consumers started to make their choice on the basis of this information and by doing so, exerted effective discipline on restaurants leading to an increase in average hygiene. Also, there is an extensive literature on (voluntary or mandatory) product labelling. Again, empirical studies tend to find a clear relationship between a change in consumer attitudes and the introduction of labelling on nutrition (e.g. Frazao and Allshouse 1996, Mathios 2000 or Loureiro et al 2006). Also, Howard (2005) shows that once the difference between expected and actual graft failure rates (quality) of transplant centres was published, patients waiting for kidney transplantation started to base their transplant centre choice on this information.[16] Similarly, Foreman and Shea (1999) finds that as statistics on on-time performance of US domestic airlines went public, consumers re-adjusted their demand and exerted effective discipline on firms leading to an improvement of average quality.

On the other hand, there has been less work on imperfect quality indicators and how consumers react to them. The results of this literature are mixed, which can perhaps be taken as evidence that consumers find it difficult to understand quality indicators. Hodgkin (1996) finds no evidence that patients in general take the availability of specialised cardiac services as a signal of quality when they choose hospitals. Mocan (2007) cannot confirm that parents are strongly rational when making choices among child care services: they often give importance to uninformative quality signals while ignoring informative ones. Reinstein (2005) finds a very small but detectable effect of movie reviews on film choices.

1.2 Empirical literature on audit quality

Audit quality is unobserved and, perhaps not surprisingly, very little empirical work has been done to uncover its role in the audit industry. Empirical studies often assume quality differences across auditors but do not test their hypotheses. For instance, it is quite common in the literature to assume that the big 4 (6,8) offer superior quality to non-big 4 (6,8) audit firms (e.g. Farber 2005, Khurana and Raman 2004, Gul et al 2002, Reed et al 2000).[17] In another study, Jiraporn (2006) shows that firms with weaker shareholder rights have a tendency to select Arthur Andersen (which is assumed to represent lower quality) over the other big six. Other articles focus on measuring the adverse effects of reputation

[16] A detailed discussion on why it is relatively easy to measure quality in transplantation can be found in Howard (2006). In short, kidneys are too precious goods to experiment with: a kidney gets transplanted into a patient if and only if the probability of success is almost certain as described by a set of well defined criteria beyond the discretion of the hospital. So patient and kidney characteristics have only a minor influence on the actual probability of transplantation success.

[17] Note that the fact that there are significant fee differences (on average around 20%) between the big 4 (6,8) and smaller accountant firms is by no means evidence for different audit quality. For instance, according to the 'insurance hypothesis' (Francis 2004), investors and shareholders prefer and, as a consequence, firms may want to choose a big auditor because of its deep pocket: in the case of a lawsuit, big auditors will more probably stand for financial damages. Also, other obvious differences between the two groups of auditors such as industry experience can explain the substantial gap between big 4 and non-big 4 fees.

damage on the market value of client companies due to auditor failure of some sort (e.g. Krishnamurthy et al 2006, Pacini and Hillison 2003, Chaney and Philipich 2002).

However, there are very few studies investigating auditor quality directly. Fuerman (2003) tests if there is significant difference in quality (proxied by auditor litigation history) between big 5 auditors individually and finds that all big 5 auditors but Arthur Andersen offered superior quality compared to boutique (non-big 5) auditors. However, Tilis (2005) investigates along a longer period and finds no quality differences among big 5 (6,8) and across big 5 (6,8) and boutique auditors. Her quality measure is built on the idea that if the standard deviation of a company's abnormal accruals is high, then it should be associated with large errors in its financial reports, and therefore should indicate low audit quality.

To my knowledge, however, audit quality *perception* among clients and how these differences in perceived quality actually affect clients' auditor choices has not been analysed. As a result, we don't know how potential clients perceive quality and whether they act upon their perceptions. In other words, we seem to know little about how effective and disciplinary the reputational mechanism is in the audit industry. In particular, it seems to be well documented how capital markets react to quality signals but there appears to be no empirical work on how this capital market reaction feeds into companies' auditor choice. This is the scope of the current study.

This study is, therefore, not concerned with measuring actual quality differences. Rather, it tries to understand quality perceptions across potential audit clients. In particular, I am interested in how readily available market information affects auditor choice, that is, whether the reputational mechanism is effectively at work or not. Companies usually keep a close eye on competitors so it is reasonable to assume that they are aware of a restatement announcement of their peers and its market effect. Since restatements have a direct effect on companies' market value, it should be an important indicator for them and as such should influence their choice to a detectable extent.

In the current framework it is possible to test for differences in quality perceptions among big 4 as well as between big 4 and non-big 4 public accountants. The results suggest that there exist significant differences in perceived quality within the big 4 and also between the big 4 and non-big 4 auditors. Therefore, the conventional quality classification in the audit literature (big 4 – high, non-big 4 – low) appears overly simplistic.

1.3 The audit industry

The audit industry came into existence in the USA as a result of the Securities Acts of 1933 and 1934. Establishing the principle of full disclosure, these federal securities laws obliged public companies to provide investors with full and accurate information and required them to have their financial statements audited by independent accountants. The public company must secure an independent audit opinion on a yearly basis in order to be listed on a national stock exchange and to comply with reporting requirements of the securities laws. Bringing integrity, objectivity and professional competence, the external auditor is, therefore, to ensure that a company's financial statements are fairly presented and reliable in all material respects. In other words, the public accountant serves as an independent link between the management and the investing community.

At the beginning, competition among auditors was severely impaired: advertising, competitive bidding and soliciting clients were all strictly prohibited. As a result, hundreds of audit firms operated undisturbed for three decades. Beginning from the 70s, however, rules restricting auditors from advertising and competitive bidding were loosened unleashing fierce competition. (GAO-03-864) Ever since, the industry has been going through a remarkable concentration process. By the 1980s, eight firms dominated the American audit market and by 1998, this was down to five as a result of a series of mergers among the big 8. Then, the collapse of Arthur Andersen further reduced the number of dominant firms to four. Whereas only 22% of the smaller companies (revenue less than $100 million) were audited by the big 4, the large company market is very concentrated: 96% of the companies with revenues over $500 million were big 4 clients in 2006.[18] (GAO-08-163)

Public accountants used to offer a wide range of services. In addition to audit, they developed substantial non-audit, typically consulting arms. These included but were not limited to tax, information technology and management consulting. Non-audit business lines grew into a substantial source of income over the years: by 1998, management consultancy alone on average constituted 45% of the revenues of the big 5. (GAO-08-163) These services, however, seriously called into question auditor independence and therefore public accountants started to divest or sell these business lines from 2000. In 2002 the provision of a wide range of non-audit services to audit clients was explicitly prohibited in the Sarbanes-Oxley Act.

The market for audit services is typical of an industry where moral hazard plays a major role. Auditors verify financial statements every year to ensure that investors get a true picture of the company. Potential investors and shareholders must rely on a credence claim: the veracity and authenticity (quality) of an audit is never directly observed. Public companies, however, have a strong incentive to misreport their financials and, as a result, so do auditors who are paid by them. On the other hand, investors are well aware of this incentive and therefore, they want to learn about auditor conduct in every possible way. If an auditor's reputation suffers serious damage, investors cease to rely on the financial statements it has audited and the market confidence in the value of its clients evaporates. Therefore, although the company has a quite natural driver to misreport its financials, it also has a strong incentive "to appear" to report truthfully. As a result, potential audit clients should care about auditor quality. Although audit quality

[18]Market shares based on audit fees as of 2006: PWC 30.3%, EY 23.5%, DT 21.2%, KPMG 19.4%.

is not observed, there is information available that market participants may regard as quality indicators. The auditors' financial restatement history could well be such an indicator.[19]

1.4 Background: the Enron scandal

Arthur Andersen audited the Texas based energy company from its birth in 1986. In 15 years, Enron grew into the US's seventh largest firm. Through the years, Enron became one of Andersen's biggest clients worldwide paying tens of millions of dollars year after year for audit as well as consulting services. Although the fraudulent accounting practices are thought to have started as early as 1996, market confidence in the company began to fade only in 2001.

On October 17, 2001 the Enron scandal started to unfold: the chairman, Kenneth Lay announced that the third quarter results were negative – something that had been unheard of before. This reinforced fears about the inadequacy of Enron's financial records, which first had seriously emerged on the 15 August, when Jeffery Skilling, CEO, the main architect of the company's success, unexpectedly resigned for personal reasons. On October 23rd, the SEC announced that it was investigating the company. Enron filed for bankruptcy on December 2, 2001. Arthur Andersen admitted shredding documents on January 10, 2002 and its criminal indictment was announced on March 14, 2002. On 15 June, 2002 Arthur Andersen was convicted of obstruction of justice and the firm surrendered its licenses on August 31, 2002. As a result, approximately 1600 companies were forced to choose a new auditor between October 2001 and August 2002. 87% of these companies chose big 4 auditors. Although auditors had gone bankrupt before Andersen's demise,[20] Arthur Andersen' failure was the single most important event which seriously called into question audit quality and ultimately the very existence of the industry.

It's hard to pin down when Arthur Andersen's conduct started to be seriously called into question and, as a result, companies started to dismiss it. I take October 1, 2001 as the start date of the accounting (Enron) scandal, similarly to the study of the US Government Accountability Office (GAO) on Public Accounting Firms (GAO-03-864).

2 Data

The sample consists of a cross section of 1262 US public companies which switched from Arthur Andersen between October 1, 2001 and August 31, 2002. The data come from several sources. Audit fees and information on the new engagements of former Arthur Andersen clients come from AuditAnalytics. Since

[19] Financial restatements are, of course, not the only indicators. For instance, litigation history of an auditor could also be a straightforward and natural measure. However, primarily for data availability, this paper analyses only the effect of financial restatements on auditor choice.

[20] For instance, one of the leading cases was Laventhol & Howarth, the seventh biggest auditor at the time, which in 1990 filed bankruptcy as a result of a row of accounting scandals.

this source has very limited information on company financials, I also gathered balance sheet and income statement items from Compustat to supplement AuditAnalytics data. Unfortunately, the vital financial information on companies is deficient even after supplementing AuditAnalytics with other sources as is clear from the Summary Statistics in Table 1. I use several variables to control for auditor and client characteristics that are described in details in the following sections.

2.1 Auditor Characteristics

I control for auditor characteristics along three dimensions: auditors' restatement history (quality), industry experience and audit fees.

Financial Restatements

The key variable of interest is financial restatement frequency. Although SEC and policy makers in general consider them as audit errors, the extent to which financial restatements can be taken as audit failures is debatable. Firstly, capital market reactions seem to differentiate among restatements that vary along numerous dimensions. For instance, restatements can be initiated by the SEC, the company, the auditor or investors. Also, the report to be restated can be quarterly or annual. In addition, the reasons for restating can be several.[21] Timing can also be important: as the restatement period increases, audit quality is growing more and more questionable. The stock market reaction to restatements appears to vary along these characteristics as well as time. For instance, before Andersen's collapse, on average revenue recognitions triggered the most severe and prolonged fall in share prices of the restating companies,[22] whereas post-Enron cost/expense restatements had the most adverse effect (GAO-03-138, GAO-06-678). Secondly, it is not clear to what extent the auditor is responsible for a restatement. For example, quarterly reports, though reviewed and signed off by the auditor, are not audited so the auditor has less responsibility for possible misstatements. On the other hand, between 1995-1999 auditors were sued for quarterly restatements (4% of restatements), although much less often than for annual ones (23%). (Palmrose and Scholz 2004)

Even though restatements differ a great deal, in this paper I do not differentiate among them. The reasons are twofold. Firstly, the data I have do not allow me to classify restatements along many interesting features. And even if they did, there is no common agreement as to how to classify restatements. Probably though, it would be beneficial to differentiate among restatements at least in one important respect: the auditor's responsibility clearly differs if the restated report is annual (audited) or quarterly

[21] GAO classified restatements into nine groups: M&A, Cost or Expense, R&D, Reclassification, Related Party transactions, Restructuring & Inventory, Revenue Recognition, Securities Related, Other (GAO-03-138)

[22] Though accounting only for 39% of the restatements, revenue recognition resulted in 56% of the total capital loss.

(non-audited). Unfortunately, none of the datasets recorded this (or other relevant) feature. Secondly, although there has been no common way to classify restatements in the literature, the trends and conclusions are very similar so there seems to be little gain from a tedious classification. (GAO-03-138) This is confirmed by my data. I have two data sources for restatements which were collected in very different ways (see below for details). The restatement variables created from the two different data sources are strongly correlated (0.97) and performed very similarly in the choice models as we will see shortly.

As mentioned above there are two possible sources for variables of financial restatement frequency. In its study of Financial Statement Restatements (GAO-03-138), GAO identified 919 restatements between January 1, 1997 and June 31, 2002, which were the results of erroneous and/or fraudulent accounting practises and therefore can be safely taken as audit errors. Unfortunately, the GAO database was not sufficiently detailed to identify the industries of restating companies, moreover, they didn't record any suitable company identifier (e.g. CIK code) by which I could identify companies' industries in other databases. Therefore I compiled an alternative dataset on restatements from AuditAnalytics. The AuditAnalytics restatement data, however, are only for 2001 and 2002. It differs from the GAO data in another respect too: it contains all financial restatements rather than only those when erroneous and/or fraudulent accounting is believed to be behind the restatement.

Therefore, I will construct different measures for restatement frequency using two different datasets. First, with the aid of Who Audits America (1997-2001), I identified the auditors of the restating companies in the GAO database and then for each auditor I calculated the percentage of audit clients which restated their financial statements in this period and so created *Rest. (A.,GAO)*. I also calculated the very same variable from the AuditAnalytics database, *Rest. (A.,AA)*. However, these measures will not be sufficient in the choice models of both audit and client characteristics because these two restatement variables vary only with auditors and, as a result, they would be perfectly collinear with auditor fixed effects. Therefore, I had to calculate alternative restatement variables which vary along other dimensions. In principle, there could be two possible sources of additional variation: perceived restatement frequencies can also vary across industry and time.

Variation across industries comes from clients' different industrial background. It is hypothesised that clients are not particularly concerned with the overall restatement history of the auditor but they rather focus on the restatements which occurred within their own industry since it is the events of their peers that they follow most closely. So a bank, for instance, knows and cares mostly about an auditor's restatement history in the financial sector as opposed to its overall restatement performance. Therefore, I calculated restatement frequencies by auditor and industry: *Rest. (A.I.)* will be my main quality variable. Note again this variable can only be computed from AuditAnalytics since GAO didn't identify the industries of the restating companies.

Time variation of perceived restatement frequencies comes from the fact that clients broke ties with Arthur Andersen at different points of time. Depending on which month a client dismissed Andersen and chose a new auditor, each client would have different information on the proportion of an auditor's restating companies resulting in differing perceived restatement histories across clients. This source of variation doesn't seem to be of much promise, however, because the vast majority of clients left Arthur Andersen between March and July 2002, offering little hope for sufficient variation in restatement histories across time. Also, one may object that it's difficult to pin down when an actual decision of the company is made on the engagement of the new auditor. Nevertheless, *Rest. (A.I.T.)* varies with auditor, industry and time and is calculated from AuditAnalytics again.

Industry Experience

Auditors' industry experience can be measured in different ways. A crude measure could be the number of clients of an auditor in a given industry. I obtained this data using AuditAnalytics' Audit Opinion. According to this dataset, in 2001 the big 4 audited 7774 and the non-big 4 auditors 1674 public companies. The *Ind. Exp. (N)* variable is the total number of companies in the potential client's industry audited in 2001 for each auditor.

A more sophisticated measure is the variable *Ind. Exp. (MS)*. Here, I calculated what percent of the market (total assets) of a given industry is audited by each auditor. Then I matched the results again with the client's industry. So, for instance, for a bank, I calculated the total market share of the companies in the financial sector each auditor audited in 2001. In the choice models this measure will be my main industry experience variable.

Audit Fees

Audit fees are paid for audit related services and are reported annually. Companies had to disclose this information only from 2000 in the USA and these data have been collected from SEC files by AuditAnalytics. A potential problem is that I have data only on fees actually paid so I don't observe a company's virtual fee that it would have paid, had it engaged an alternative auditor. I will, therefore, need to estimate audit fees for each company (for more details see below).

Non-audit fees are payments that a company incurs for other than audit, typically consulting services. Non-audit fees are excluded from the analysis for two reasons. First, as mentioned in the introduction these fees played a limited role by 2002 because from 2000 public accountants started to divest and sell off these business lines as a response to the growing general perception of the auditors' impaired independence.[23] In 2002, however, with the implementation of Sarbanes-Oxley Act, public accountants were explicitly banned from providing a wide range of non-audit services to their audit clients. Furthermore, it is not clear to what extent audit and non-audit services were linked in 2002 (if they ever were at all). Perhaps a good example for this is the fact that Arthur Andersen did not fail as a whole: its consulting division was reorganised under the name of Accenture and has been extremely successful ever since. This alone suggests that consulting services were most probably perceived as separate businesses. Second, AuditAnalytics non-audit fee data does not contain important items: it includes tax and IT but excludes management consulting fees. Since I couldn't properly account for the entire range of non-audit service fees, they didn't seem to play an important role anyway and their link to the audit market is questionable, I therefore excluded them in what follows.

A summary of the variables with expected signs (effect on auditor choice) in brackets is as follows:

- Rest. (A.I.) (-): restatement frequency by industry calculated over January 1, 2001 – August 31, 2002 using AuditAnalytics database. This variable varies with auditor and industry.
- Rest. (A.I.T.) (-): restatement frequency by industry and time calculated over January 1, 2001 – August 31, 2002 using AuditAnalytics database. This variable varies with auditor, industry and time.
- Rest. (A.,AA) (-): total restatement frequency calculated over January 1, 2001 – August 31, 2002 using AuditAnalytics database. This variable varies only with auditor.
- Rest. (A.,GAO) (-): total restatement frequency calculated over January 1, 1997 – December 31, 2001 using GAO and Who Audits America datasets. This variable varies only with auditor.
- Ind. Exp. (MS) (+): the market share (total assets) in a given industry an auditor audited in 2001
- Ind. Exp. (N) (+): number of clients in a given industry an auditor audited in 2001
- Audit Fee (-): estimated as a function of client characteristics (for more details see below)

The summary statistics of the variables can be found in Table 1.

2.2 Client characteristics

Client characteristics include variables which capture the client's financial profile (size, profitability, risk), its industry background and several other dimensions. Most of the client characteristics will enter both

[23] This is confirmed by my data: the correlation between audit and total (audit plus non-audit) fees is 0.986.

the audit fee regressions and the choice models. In the audit fee regression, the sign of the effect of risk is positive a priori since a riskier client increases the probability of litigation against the auditor. Also, the bigger the company, the more complex the audit is, so size and complexity are expected to have positive effects too on audit fees. Lastly, profitability is thought to have negative effect, since a profitable company is less likely to go bankrupt and again means smaller litigation risk. Also, a successful company has less incentive to deceive the auditor, which implies an easier audit process. However, the a priori sign of these client characteristics is more ambiguous in the choice models. Most of these measures are calculated from the companies' financial statements and come from both AuditAnalytics and Compustat.

I also have data on SEC investigation, internal control problems, etc., which are potentially very useful risk measures. These essential data were disclosed in SEC files at the time of the auditor change and can be found in the AuditAnalytics database again. However, for most of the variables there were only very few observations available and, consequently, in the regressions they performed very poorly. Therefore, in what follows I will exclude these variables with one exception. If the auditor thinks that the company is on the verge of bankruptcy, then it qualifies its audit opinion as 'going concern'. The variable *Goingconcern*, therefore, is a classical risk measure of the client company: it takes the value one if the company received a qualified audit opinion previously and zero otherwise.

The variables are listed below with the expected signs on audit fee in brackets, followed by the definition and the effect they are to capture:

Size and Complexity

- Assets (+): in 1000 million dollars

- Revenue (+): in 1000 million dollars

- Market Value (+): total shares outstanding multiplied by stock price in 1000 million dollars, company value

- Inventory (+): in million dollars

- Business segments (+): number of business segments (corresponds roughly to product lines/divisions), measure for company complexity

Profitability

- Loss (+): a dummy taking the value 1 if the company produced a loss in the previous year

- ROA (-): net income / total assets

Risk

- CATA (+): current assets / total assets, a liquidity measure
- Inventory/Revenue (+): audit risk
- Quick Ratio (-): current assets (less inventories) / current liabilities, liquidity measure
- Goingconcern (+): 1 if the company previously received an audit opinion qualified with going concern
- Leverage (+): total liabilities / total assets, leverage

Other variables

- Dec (+): a dummy taking the value 1 if the company's financial year ends in December
- Industry dummies (+/-): agriculture, natural resources, utilities, manufacturing/construction, trade (retail, wholesale), services, information technology, financial sector; all based on NAICS codes
- Dummy Zero (+/-): dummies for zero values of variables

For the summary statistics, please see Table 1. Most of the variables are self-explanatory. *Inventory/Revenue* is included because inventory is notoriously difficult area to audit and often subject of restatements. Therefore they increase the risk of an audit client. *Dec* dummy is an instrument (see for details below). I also included dummies for zero values whenever I took the log of a variable. In order to avoid missing values, I rescaled the variables by adding 1 to each observation. However, this method is sensitive to the rescaling value, so I decided to include a dummy for the observations when the original variable takes the value of zero.

3 Empirical strategy

First, I will estimate simple multinomial logit models in which only auditor characteristics explain auditor choice. Next, I will present results of multinomial models where the choice is a function of both auditor and client characteristics to see if client specifics play a role in determining choice. Of course, these models are very restrictive in terms of the substitution patterns that they allow for. To relax these restrictions at least to some extent, I will also estimate nested logit models.

3.1 Estimation procedures

In order to estimate the auditor choice models first I need to estimate audit fees. In discrete choice models, it is imperative to observe characteristics for each alternative. However, in my data only those audit fees are available that the client did pay to its newly engaged auditor. The audit fees that it would have paid to other auditors, the virtual price of alternative audit services, are not observed; therefore they need to be estimated. Moreover, audit fees are considered to be endogenous and I use instruments to correct for this.

Thus, first I regress audit fees on different client characteristics including instruments over the whole sample. Once the best model is identified (see below), I estimate that model for each auditor separately. These regressions will be my auditor specific "pricing equations". Then I use the fitted values of these regressions to generate audit fees for each alternative for each client. These audit fees along with other variables enter the second stage choice models as a regressor, which results in consistent estimates (Rivers and Vuong 1988, Wooldridge 2007).

In what follows I briefly review the choice models to be estimated and outline the bootstrap procedure I use to correct for the standard errors.

3.1.1 Multinomial Logit

By choosing auditor j, company i gets utility

$$U_{ij} = W_{ij} + \varepsilon_{ij} \tag{1}$$

where W_{ij} is observable up to some parameters and ε_{ij} is unobserved by the econometrician. Then, company i chooses auditor j with probability

$$P_{ij} = \Pr(W_{ij} + \varepsilon_{ij} > W_{il} + \varepsilon_{il}, \forall l \neq j)$$

If ε_{ij} are independently and identically distributed type I extreme value (Gumbel), then after some algebraic manipulations

$$P_{ij} = \frac{e^{W_{ij}}}{\sum_j e^{W_{ij}}}$$

The multinomial logit models are based on the assumption that the error terms in equation (1) are identically and independently distributed (IID). This assumption has a behavioural association with the well known property of Independence of Irrelevant Alternatives (IIA). The IIA assumption may hold as a result of correct specification: if all characteristics and attributes are controlled for in the estimation, then only white noise is left in the error (Train 2002). The IID (IIA) assumption in most applications, however, are too strong since there might be factors affecting the choice that are not specified in the model. If these unobserved characteristics share some common components across observations, it will induce correlation in the error terms of those observations. Therefore, a logit model which allows correlation within a subset of choices (nest) relaxes the IIA assumption to some extent. This is the nested logit model.

3.1.2 Nested Logit

In the nested logit specification, the components in (1) can be broken down into nest and alternative specific parts. In particular,

$$U_{ij} = \underbrace{V_{iB} + V_{ij|B}}_{W_{ij}} + \underbrace{v_{iB} + v_{ij|B}}_{\varepsilon_{ij}}$$

where V_{iB} depends only on variables which describe nest B and the variables in $V_{ij|B}$ describe alternative j in nest B. Similarly, the random component in (1) can be decomposed into two parts: v_{iB} is the unobserved (stochastic) variation across the nests and $v_{ij|B}$ is the random variation within the nest. Note that the presence of v_{iB} induces correlation across the errors within the nest relaxing the assumption of IID errors in (1) to some extent. However, across the nests errors are assumed to be independent. Then the choice probabilities can be expressed as

$$\begin{aligned} P_{ij} &= P_{iB} \cdot P_{ij|B} \\ &= \frac{e^{V_{iB}+\frac{1}{\lambda_B} IV_{iB}}}{\sum_B e^{V_{iB}+\frac{1}{\lambda_B} IV_{iB}}} \cdot \frac{e^{\lambda_B V_{ij|B}}}{\sum_B e^{\lambda_B V_{ij|B}}} \qquad \text{where} \\ IV_{iB} &= \ln \sum\nolimits_{j \in B} e^{\lambda_B V_{ij|B}} \end{aligned} \tag{2}$$

IV_{iB} is the Inclusive Value or Logsum, the expected utility of choosing an alternative in nest B. In other words, it is a weighted average of the utilities associated with the alternatives within the nest, where the weights are probabilities of choosing each alternative. As it is clear from (2), the probability of an alternative being chosen can be broken down into two parts in a nested framework: it is the product of the probability of choosing the nest and the conditional probability of choosing an alternative within the nest. These two probabilities correspond to two logit models, one for the upper ("choice between nests") and one for the lower level ("choice within nest"). The two logit models can be estimated separately (sequentially) or jointly. The former is the Limited Information Maximum Likelihood (LIML) and the latter is the Full Information Maximum Likelihood (FIML) approach. Both approaches have advantages and disadvantages. Although FIML is more efficient, the numerical maximisation is much more complicated due to the fact that the joint maximum likelihood function is not globally concave. This may result in an abundance of local maxima and in some applications this can severely affect the results. Convergence may not be achieved and even when achieved the convergence point might be of dubious value. For instance, Koppelman and Bhat (2006) report models when two very similar sets of initial values resulted in the algorithm converging to two very different parameter vectors.

Although the computational problems of FIML do not emerge in many applications, they turned out to be quite an issue in the present analysis for the following reasons. The industry structure clearly partitions the choice set into the groups of big 4 and non-big 4 accountants, therefore, the nested logit models are estimated with two (Big, NonBig) nests. However, the NonBig nest is degenerate since non-big 4 accountants are represented as one single choice as a result of insufficient number of observations. Furthermore, this degenerate nest has very few observations compared to the Big nest but even compared to an alternative within the Big nest. As a result the models often failed to converge. Moreover, even if they did, they converged only after couple of thousands of iterations and usually returned meaningless results in the sense that nothing was (individually or jointly) significant.[24] To make sure, it is indeed the degenerate nest which gives me the grief, I re-estimated the models with different nesting structures where both nests contained at least two alternatives. That is, I put one or two big 4 auditors in addition to the non-big alternative in the NonBig nest. Perhaps not surprisingly, the convergence problems disappeared and the models converged after around a hundred iterations, which is as expected given the number of observations and coefficients estimated. While the results clearly suggested that the nesting structures were wrong (the inclusive value (IV) parameters were outside the unit interval in all cases, although often not significantly), yet the estimated parameters of interest often confirmed the previous results: in most of the models the restatement and industry experience variables were significant at least at the 5% level with the expected signs. The significance of the audit fee variable varied from model to model but its sign was robustly negative.[25]

For these reasons, the nested logit models are estimated sequentially. Although it's less efficient, the numerical maximisation in the sequential estimation is straightforward since the likelihood function of the multinomial logit is globally concave. As a result, this estimation procedure can be conveniently bootstrapped. There is a delicate issue, however, of normalising the sequential estimator for consistency with random utility maximisation. This problem is discussed at length in the Appendix.

[24] Note that regardless of convergence, the number of iterations in and of itself renders FIML computationally infeasible in the current context given that I need to bootstrap the results.

[25] The estimations are not reported for two reasons. Firstly, and most importantly the main results are very similar to the results reported. Secondly, if I take the estimations seriously, then the IV parameters are not statistically distinguishable from one (especially after bootstrapping), therefore the IIA assumption holds and the simple multinomial logit specifications in the previous sections are appropriate.

3.1.3 Bootstrap

Since I have generated regressors and I also correct for the endogeneity of audit fees, the standard errors are, of course, incorrect in the second stage. Instead of deriving and calculating the correct asymptotic variances, however, it is convenient to bootstrap and therefore in all choice models where audit fees appear as a regressor I use nonparametric bootstrapped standard errors with 1000 replications. The bootstrap procedure applied is as follows. In each simulation, I take a bootstrap sample from each client set that chose a particular auditor and using this sample I estimate the audit fee regressions for each auditor ("pricing equations"). Then, using these regressions I generate audit fees for all clients in the whole sample. Then I take a bootstrap sample again and run the choice model. This procedure is repeated 1000 times. Since audit fees get regenerated in each simulation step, the first stage variation feeds into the second stage resulting in asymptotically correct standard errors.

3.2 Identification

In all choice models estimated in the current study identification relies on exclusion restrictions which take the form of instruments affecting audit fees but not auditor choice. *Dec* dummy which takes the value one if the company's financial year ends in December is my primary instrument. The month the financial year ends in supposedly affects audit fees because of the auditor's uneven workload during the year. In particular, for the vast majority of companies the financial year ends in December putting auditors under strong pressure in the period of January - March, while other months of the year are usually relatively quiet. Therefore, if a company requires audit out of "rush months", it is supposed to get a discount.

However, the end of financial year is clearly exogenous to the choice. In other words, when exactly during the year the company requires an audit does not affect directly which auditor the company chooses, only through audit fees. It is hard to imagine a casual relationship between choice and the end of the company's financial year which would render the instrument invalid. The only way I can possibly think of may come about through capacity constraints. Under the circumstance of severe capacity constraint, an auditor might be unable to take on a client which would further strain its busy period, while it could easily accommodate the company with financial year ending, say, in June.[26] However, two important observations work against this argument. First, in any case this argument would be valid only in the short run. Furthermore, the "audit technology" is predominantly labour intensive and as such capacity is easily adjusted even in the short run. Second, most of the clients switched from Andersen in the first half of the year leaving ample amount of time for auditors to adjust their capacity for the coming "audit

[26]Note that this argument would also defeat the whole modelling concept since it is assumed that companies choose auditors and not vica versa.

peak" at the beginning of the next year had they needed to.[27] As a consequence, we can safely assume that factor (labour) demand was not completely inelastic, that is, there existed a price (audit fee) for which any auditor would have been willing and able to increase its capacity. In addition, it is clear from Table 4 that each big 4 auditor and also non-big 4 auditors engaged companies of both types (i.e. companies with financial year ending in December and in other months) which again suggests against the idea of binding capacity constraints. But this just means that the end of financial year affects choice only through audit fees. Therefore, I regard my instrument as strongly exogenous.

I also create another instrument interacting Dec dummy with Assets. This in principle opens the door to overidentification tests. However, overidentification tests rely on the assumption that at least one of the instruments is exogenous. In my case, exogeneity comes from the same source: if one of the instruments is endogenous so is the other. Nevertheless, I used a commonly applied overidentification test for the sake of completeness (Murray 2006). I used one instrument to instrument Audit Fees and then included the other instrument as an explanatory variable in the choice model. If both instruments are truly exogenous, then the instrument as an explanatory variable in the choice model should be insignificant. This is indeed the case: the resulting p-value is 0.71.

The joint F statistic of the instruments in the regressions that I eventually use to generate the audit fees is 9.23. It is possible to include further (insignificant) variables in the audit fee regression and push the F statistic slightly above 10, the commonly used cut off value of weak instruments. However, as discussed below the final results are very robust to the specification of audit fee regressions. Moreover, it is not obvious how relevant the threshold of 10 is in non-linear settings. But in any case, there is little doubt about the strong exogeneity of the instruments which effectively eases the problem of weak instruments.

4 Results

Throughout the analysis I will have 4 choices (big 4 auditors) plus one, the non-big 4 auditors together as the fifth alternative. Unfortunately, I am unable to further refine the choice set. In the multinomial models, I report results on two sets of choices. In the first one, the models are run on the whole sample where the five alternatives are PWC, EY, DT, KPMG and NonBig, the latter being non-big 4 public accountants as a fifth single alternative. In the second, I only include the choices of the big 4 accountants. As we'll see shortly, in terms of the final results, the two sets of estimations arrive at very similar conclusions. Later in the nested logit framework I explicitly make the distinction between the two

[27] One may object saying trying to re-establish the integrity of their financials, some companies required an immediate audit on previous years' financial statements. Note that, however, since most of the auditor changes took place in the first half of the year, an immediate audit meant for the auditor "work out of rush months" implying unlikely capacity problems.

groups of big 4 and non-big 4 auditors.

However, before discussing the results of the choice models, let me review briefly the estimations of the audit fee regressions.

4.1 Audit fee regressions

In the first stage I estimate audit fees in several regressions of the following generic form:

$$\log(Audit\ Fees_i) = \alpha + \beta_1' \log(Size_i) + \beta_2 Complexity_i + \beta_3' Profitability_i + \beta_4' Risk_i + \theta' Instruments_i + \varepsilon_i$$

The audit fee regressions for all (big 4 and non-big 4) clients can be found in Table 2, whereas the same regressions for big 4 clients only are in Table 3. As it is obvious from the tables of both sets of observations, client characteristics seem to robustly predict auditor fees. 74% of the variation in the dependent variable is explained by the final models, which is quite standard in the audit fee literature. In addition, all the variables have the right signs and most of them are significant at the 1% level consistently across all models reported. Perhaps surprisingly, both the significance and the magnitude of the parameters appear very similar in the two tables. Note that around 10% of the observations are non-big 4 clients, which would be a sufficient fraction to change parameter values should pricing differ a great deal across big 4 and non-big 4 auditors.

I will use the results from these regressions to generate audit fees for all auditors.[28] Some notes on these regressions are in order. Since the predictive regressions are often run on a small sample (auditor by auditor), it appeared important to save degrees of freedom. As a consequence, I only included variables which have fairly accurate predictive power. In particular, I included a couple of measures for size, company complexity, risk and profitability and then I dropped variables in each category which were insignificant at the five percent level. In other words, one can find more than one variable within a category (e.g. size) only if they are all strongly significant. The final results are robust to this modelling decision: including different sets of insignificant variables ($p > 0.05$) in the first stage changes only the significance of the audit fee variable in the second stage, while other second stage coefficients are basically unaffected. For instance, leaving insignificant industry dummies in the audit fee regressions leads to a significant reduction in the variation of the predicted audit fees. The reason for this is that (insignificant) industry dummies mop up within industry variation so the regressions will predict very similar audit

[28] One may argue that audit and non-audit fees are related in a way that high audit fees may be accompanied by lower non-audit fees or vice versa. So I reran the estimations with total (audit plus non-audit) fees. The results didn't change which is no wonder since the correlation between audit fees and total fees is 0.986 in my sample.

fees within an industry removing a significant source of variation. The list of variables used for audit fee predictions as well as the estimation results can be found in Table 4. I estimated the "pricing equations" separately for the five alternatives and calculated the fitted values of these regressions for each client. In this way, for each company I obtained the estimated audit fees for all alternatives. The summary result of these estimations can be found in the summary statistics in Table 1 under the heading of Auditor Characteristics.

Furthermore, note that in the audit fee regressions in Table 2 and 3 and also in the choice models later, I include inventory as a size (Inventory (log)) and also as a risk measure (Inventory/Revenue). This may make little sense since I already use quite a few variables to control for size so the variable Inventory (log) seems unnecessary. However, note that Inventory (log) is strongly significant in the audit fee regressions, therefore, I did not want to carelessly discard it. In addition, since it's strongly significant in the audit fee regressions, I see no compelling reasons why I should exclude it in the choice models a priori. Since the correlation between Inventory (log) and Inventory/Revenue is quite low (0.11), I see no harm in the inclusion of both for the sake of complete transparency. However, I have experimented with three different specifications: a) only Inventory/Revenue in all (audit fee and choice) models, b) only Inventory (log) in all models, c) Inventory (log) in audit fee and Inventory/Revenue in choice models. The main results were completely invariant.

4.2 Multinomial Logit models of Auditor Characteristics

In this section, the specification of (1) is of the general form:

$$W_{ij} = \eta_1 Restatements_{ij} + \eta_2 Industry\ Experience_{ij} + \eta_3 \overline{Audit\ Fees_{ij}}$$

where, *Restatements* and *Industry Experience* is one of the restatements and industry experience variables respectively, defined in section 2.2 and $\overline{Audit\ Fees}$ is a linear projection from the first stage estimations. Again, I report two sets of results: in Table 5 big 4 auditors separately and non-big 4 auditors as a fifth single alternative are the available choices, whereas in Table 6 only big 4 accountants are included in the choice set. As mentioned earlier, the standard errors have to be corrected, therefore all standard errors in the tables are (nonparametric) bootstrapped with 1000 repetition. The results differ a great deal along the two sets of specifications in some respects so I discuss them separately.

In the estimations with only big 4 auditors, the results show a fairly consistent picture: all variables are strongly significant with the expected signs. It can easily be seen that the restatement variables have a robust effect: across all specifications they are significant at the 1% level with the expected negative sign. That is, more financial restatements decrease the probability of being chosen. Note that this, along with the results in Table 8 (see below), means that there are significant differences in perceived quality within the exclusive group of big 4 auditors. This finding is in sharp contrast to the commonly held view of quality in the audit literature (e.g. Francis 2004) and suggests that the usual quality classification in the audit literature (big 4 – high, non-big 4 – low) is overly simplistic. In this sample the results are robust to the definition and construction of the restatement variables.

Also, both industry experience variables have a positive and (mostly) significant effect on the probability of engagement. In addition, audit fee negatively and significantly affects the choice probability, as expected. As mentioned before, this result is somewhat sensitive to the specification of the first stage regression: the inclusion of non- (or weakly) significant variables in the first stage often renders audit fees insignificant in the second stage. However, its negative sign is robustly preserved in all cases and other variables are unaffected.

The estimation results for the whole sample (big 4 as well as non-big 4) are less clear cut (Table 5). Here, the restatement variables are insignificant at conventional levels, although they have the right sign in the first four models. Industry experience, however, seems to have a fairly robust effect in six out of eight models. Audit fees are not significant at all, moreover they have the wrong sign. However, most probably these estimations are seriously misspecified. The reason is that there is a clear difference in the characteristics of big 4 and non-big 4 clients (e.g. size). These models, however, fail to take into account these effects explicitly since client characteristics are not included. As we will see shortly when client characteristics are properly controlled for we will get a very different picture.

In what follows, I will present models only with Rest. (A.I.) and Ind. Exp. (MS) variables. The reasons are as follows. Two variables do not vary with industry (Rest. (A.,AA) and Rest. (A.,GAO)), therefore they cannot be included in multinomial models of auditor and client characteristics since they would be perfectly collinear with auditor fixed effects. The other restatement variable which varies with industry as well as time (Rest. (A.I.T.)) is also excluded because the results of the models estimated with this variable were essentially the same as the ones reported. The auditor's number of clients within an industry (Ind. Exp. (N)) as an industry experience variable didn't perform very convincingly statistically and also it is a very crude measure intuitively. Hence, I report estimations of models with the Ind. Exp. (MS) variable only, which is based on market shares.

4.3 Multinomial Logit models of Auditor and Client Characteristics

As the results in the previous section suggest it is probably important to control for client characteristics. Therefore, the following models are estimated with both auditor and client characteristics in order to see if client specifics play a role in the choice models. In this section, the specification of (1) is of the general form:

$$\begin{aligned} W_{ij} &= \mu_j + \eta_1 Rest.\ (A.I.)_{ij} + \eta_2 Ind.\ Exp.\ (MS)_{ij} + \eta_3 \overline{Audit\ Fees_{ij}} \\ &\quad + \delta'_{1,j} Size_i + \delta_{2,j} Complexity_i + \delta'_{3,j} Profitability_i + \delta'_{4,j} Risk_i \end{aligned}$$

For the results see Table 7, 8 and 9. The estimated parameters differ somewhat across the two sets of choices in Tables 7 and 8. The interesting thing to observe, however, is that after controlling for client attributes, auditor characteristics have almost identical effects in both sets of specifications in contrast to the finding of the previous section. The Rest. (A.I.) and Ind. Exp. (MS) variables are both strongly significant (mostly at the one percent level) across all models. Note that the magnitudes of the coefficients are very similar too. The significance of Audit Fees, on the other hand, seems to fade away as we introduce more and more control variables, which is natural since the identification of the parameter relies on less and less variation and ultimately only on the instruments. However, an important observation is that while the standard errors of the Audit Fee coefficients increase with the inclusion of client characteristics the point estimates appear to be very similar across all specifications. An interesting finding in column 7 of Table 7 is that the effect of the Rest. (A.I.) and Ind. Exp. (MS) variables do not differ across the big 4 and non-big 4 dimension. This is not true for the audit fees: there is evidence at the 10% level that non-big clients are more price sensitive.

Comparing client characteristics between the two sets of estimation results in Tables 7 and 8 is a bit more problematic.[29] However, we can safely draw the conclusion that client specifics do not seem to play a major role in explaining auditor choice. Perhaps surprisingly, the classical size measures like Assets, Revenue, Market Value do not seem to matter much although the variable Assets showed much more potential before bootstrapping. Insignificant size effects might be the consequence of the fact that the analysis focuses on a big 5 auditor's clientele. Moreover, the risk profile and more generally the financial position of companies do not appear to affect choice. As we have seen in the audit fee regressions, company specifics strongly determine the level of audit fees. However, they do not influence auditor choice much. Although in Tables 7 there is some evidence that unprofitable businesses did not favour EY and DT, and more complex firms (Business Segments) were less likely to choose PWC and KPMG, it is clear from the estimation results of Table 8 that these are probably auditor rather than big 4 specific findings.

[29] Choice characteristics are completely invariant to the base category, whereas individual characteristics are, of course, not.

Therefore, after controlling for auditor characteristics, we can reject the commonly held hypothesis that riskier firms chose non-big 4 auditors.

Next, let's investigate if the sensitivity to quality varies with size, risk or profitability. In Table 9 we can see how the effects of auditor characteristics vary with client attributes. Basically, with the aid of interaction terms we can break down the effect of the variables into a client-invariant and client-varying components. The main message from Table 9 is that restatements seem to be evaluated very similarly across clients: none of the restatement interaction terms is significant at any conventional level.[30] Moreover, surprisingly the effects of auditor characteristics in general do not seem to vary significantly along observable client dimensions at all. Almost all the interaction terms are insignificant at conventional levels suggesting auditor attributes are evaluated very similarly among potential clients regardless of the many distinctive client features. One exception is CATA, hence there is some evidence that riskier clients prefer auditors with less industry experience. As we saw earlier the effects of the Rest. (A.I.) and Ind. Exp. (MS) variables are strongly and robustly significant. However, in some cases, the introduction of the interaction term leads to both the client-invariant and the interaction term being insignificant. This simply means that although the total effect of the variable is strongly significant, the two components individually are not distinguishable from zero at any conventional significance level. (For instance, interacting Rest. (A.I.) with Assets) On the other hand, when they are distinguishable, then in most cases we can conveniently reject the null of any client-varying effect. (e.g. interacting Rest. (A.I.) with ROA) Nevertheless, perhaps in some cases the sign of the interaction terms can be interpreted in a meaningful way: for instance, bigger firms may care more about quality, though it is rather suggestive speculation than statistical evidence. Similarly, more profitable firms might give a greater importance to industry experience. In the case of Audit Fees the presence of size-varying effect is perhaps a bit more compelling: bigger firms seem somewhat less price sensitive.

4.4 Nested Logit

In order to relax the IIA assumption to some extent and allow for more general substitution patterns, I estimate the auditor choice models in nested logit specifications. In many empirical analyses, the nests are not obvious a priori and, therefore, an elaborate range of experiments is often necessary to identify the nest structure that best describes the data. (Koppelman and Bhat 2006) However, in the current context the industry configuration of big 4 versus non-big 4 auditors clearly lends itself to a natural partition of the choice set. Therefore, I will not dwell on exploring different nesting structures. The models reported

[30]Perhaps I should mention that before bootstrapping some of the interaction terms between Rest. (A.I.) and size measures (Assets and Market Value) were weakly significant (10%). Therefore, one may want to blame the inefficient estimation procedure for the lack of client-varying effects. In addition, as a robustness check all models were also estimated with only Rest. (A.I.) interaction terms and the results are identical to the ones in Table 9.

below are of two levels and two nests: in the big 4 nest, *Big*, are the big four accountants as separate choices and in the non-big 4 nest, *NonBig,* there are the non-big 4 accountants as a single choice. Note that given the small number of companies which chose non-big 4 auditors, unfortunately I am unable to define distinct alternatives within the NonBig nest. The non-big 4 nest, therefore, is degenerate. From (2), the general specification of the nested logit models is the following:

$$
\begin{aligned}
V_{iB} &= \gamma_1 Assets_i + \gamma_2 Leverage_i + \varphi IV_{iBig} + IV_{iNonBig} \\
V_{ij|B} &= \mu_j + \eta_{1,B} Rest.\ (A.I.)_{ij} + \eta_{2,B} Ind.\ Exp.\ (MS)_{ij} + \eta_{3,B} \overline{Audit\ Fees_{ij}} \\
&\quad + \delta'_{1,j} Size_i + \delta_{2,j} Complexity_i + \delta'_{3,j} Risk_i + \delta'_{4,j} Profitability_i
\end{aligned}
$$

Note that since the NonBig nest is degenerate its coefficient is one and it's not identified (see the Appendix for more details). As discussed earlier, the models are estimated sequentially.

It is often not obvious which variables should enter the lower and which ones the upper level. The way I proceeded is as follows. First, I estimated a model with all client characteristics entering at the lower level and having a big 4 accountant as a base category. In this way, I could identify those variables that did not vary within the big 4 nest, but their corresponding NonBig parameters were significantly different to the base category. This suggested variables which had potential variation across nests but not within the Big nest. I added then these variables to the upper level step-wise. In this way, I identified Assets and Leverage as nest-specific variables.

The estimation results can be found in Table 10. As can be seen in the table, the IV parameters are between zero and one in the last 3 models, therefore the nesting structure appears to be, by and large, appropriate. These IV parameters suggest some correlation within the Big nest although none of the IV parameters can be statistically distinguished from zero or one. Moreover, based on the likelihood values, the fit of the nested logit models is very similar to that of the multinomial logits; neither of them seems to be superior to the other.[31]

The key variables of interest, the auditor characteristics variables are very similar to the findings in the multinomial models both in terms of significance and magnitude.[32] In particular, the Rest. (A.I.)

[31] Note that all models are estimated with only partially generic variables. That is, the auditor characteristics variables are allowed to vary across partitions. The reason for this is that in the presence of a degenerate nest, it's impossible to impose the restriction of generic variables. To be more precise, it is possible to estimate models with completely generic variables (i.e. parameters which do not vary across alternatives and partitions) but then the IV parameter of the degenerate nest will relax this restriction giving the false impression that the restriction was properly imposed. In other words, the IV parameter of a degenerate nest is not identified in general. However, in the presence of generic variables, technically it is identified but has little to do with the true, underlying IV parameter. See for the details for instance Hunt (2000), Heiss (2002) and Hensher and Greene (2002).

[32] In Table 10, the parameters of the auditor characteristics corresponding to the Big nest are not rescaled by the IV

and Ind. Exp. (MS) variables are strongly significant with the expected signs. Perhaps interestingly, the Rest. (A.I.) and Ind. Exp. (MS) seem to have no variation across the big and non-big partitions, just like in the multinomial models. In contrast, audit fees yet again seem to have a significantly different impact on the probability of choice between nests suggesting more price sensitivity for non-big 4 auditor clients.

Interestingly, client characteristics do not show much potential in explaining auditor choices in the nested framework, similarly to the simple multinomial models. Curiously, size again did not seem to matter even at the upper level although it might be noteworthy that the variables at the upper level in general and Assets in particular had a much more robust effect before bootstrapping implying there is significant variation at the preceding stages feeding into the final stage.[33] Risk measures were not very successful predictors of choice between nests as well as within the big 4 nest. The former finding is perhaps a bit more surprising and goes against the commonly held hypothesis that riskier firms tend to choose non-big 4 auditors.

5 Discussion of the choice models

So far it seems that choice is basically driven by auditor characteristics. Among these characteristics financial restatement history appears to be an important quality indicator for potential clients. However, some caveats are in order.

Note that a big problem is that the Rest. (A.I.) and Ind. Exp. (MS) variables vary essentially only with industry rendering it impossible to control for unobserved industry effects with industry dummies. In principle, there should potentially be some time variation in the restatement history variable due to the fact that Andersen clients dismissed their former auditor at different points of times. However, this variation proved insufficient for identification since the vast majority of companies dismissed Andersen between March and June 2002. Consequently, one may argue that the Rest. (A.I.) and Ind. Exp. (MS) variables are merely picking up unobserved industry effects. It appears impossible to disprove this argument statistically since as it seems the information in the data has been fully exhausted. Unfortunately, there is no significant within industry variation that can be exploited to unambiguously establish the effects of these two variables.

However, bold statement as it may sound, it is difficult to imagine what industry effects I haven't already accounted for in the models of audit and client characteristics. It is very reasonable to argue that unobserved industry effects play a role in the audit fee regressions (hence, their inclusion) but it's

parameter in order to facilitate direct comparison with the multinomial results. See the Appendix for discussion on rescaling.

[33] Note that the estimation procedure actually consists of three stages here: audit fee regressions and the two levels of the nested model. Therefore, no wonder that the correct standard errors are very high.

more difficult to see what sort of industry effects which are not accounted for by the Ind. Exp. (MS) and Rest. (A.I.) variables could potentially affect auditor choice. In other words, for what reasons other than industry experience and audit quality would, say, a utility choose PWC? If there are time invariant reasons why utilities should be attracted to a particular auditor, then that effect should be picked up by the auditor's industry experience. However, in principle it is not impossible to imagine that some auditor targeted certain industries by, say, a big marketing campaign to attract clients at the time of Andersen's collapse. This effect wouldn't be picked up by the Ind. Exp. (MS) variable since that cannot potentially reflect contemporary events. Time-varying industry effects, however, are not very probable. Firstly, Andersen's sudden collapse took the whole audit market by surprise and the vast majority of clients switched auditor within two months. This short time period makes it very unlikely that any auditor could have potentially predesigned an industry tailored marketing campaign. Secondly, the audit and financial literature does not mention anything which would suggest some industry effects specifically present at the time.

One way to account for unobserved industry effects is to recreate the variables which vary only with industry using a finer industry classification than the industry dummies. As a result, the industry dummies are no longer collinear with Rest. (A.I.) and Ind. Exp. (MS). Therefore, I recalculated Rest. (A.I.) and Ind. Exp. (MS) using a finer industry classification by splitting, for instance, Trade into Retail and Wholesales Trade. Then, I reran the models of Table 7 adding the full set of (coarser) industry dummies. The results are in Table 11. Although controlling for unobserved industry effects causes the significance of the key variable of interest, Rest. (A.I.), to fade slightly, the point estimates are strikingly similar and robustly significant at the five percent level across all models. This just reinforces the argument in the previous paragraph and suggests that unobserved industry effects do not appear to confound the results of this paper.

6 Conclusion

In many markets with severe asymmetric information problems, the mechanism of reputation is supposed to discipline firms. However, this mechanism critically hinges on the supposition of strong consumer rationality. That is, reputation is an effective disciplinary force only if consumers are able and willing to decipher complex quality signals. The empirical literature failed to confirm decisively the assumption of strong rationality when consumers are individuals. In this paper, I investigate rationality in a market where clients are companies rather than people and find that they are strongly rational. Therefore, this study finds evidence to support the idea of an effective reputational mechanism in markets where buyers of the good are firms.

In particular, I analyse if former Andersen clients took auditors' financial restatement history into consideration when choosing their new public accountants. The estimation results suggest that, in sharp contrast to the common perception in the audit literature (e.g. Francis 2004), restatement performance is a very relevant and robust indicator of audit quality for companies. Therefore, the well documented adverse capital market reactions to financial restatements seem to feed into the auditor choice of clients suggesting the reputation mechanism is indeed at work in the audit market. However, this result may not apply easily to consumer choice in general. Individuals, as opposed to firms, may have more difficulty developing an understanding of complex quality signals and this could call into question the effectiveness of reputation as a disciplinary force in other markets. In addition to financial restatements, the auditor's industry experience had a robust effect. However, other variables like the price of audit (audit fees) and client characteristics in general appear to have played a very limited role in choosing auditors.

These findings have important policy implications. When consumers are found to be strongly rational, market mechanisms are effective and market crises are probably part of a solution rather than signals of its absence. The costs and benefits of direct regulation should therefore be evaluated much more carefully.

The paper also contributes to the audit literature at least in one important aspect: it is clear from the results of this study that there exist significant differences in perceived quality between big 4 and non-big 4 accountants *as well as* within the group of big 4 auditors. Therefore, the conventional quality classification in the literature (big 4 – high, non-big 4 – low) appear overly simplistic. Furthermore, I found no compelling evidence that riskier companies tend to choose non-big 4 auditors.

Table 1: **Summary Statistics**

CLIENT CHARACTERISTICS

Variable	Obs	Mean	Std. Dev.	Min	Max
Assets (MM$)	978	3259	22535	0	641100
Revenue (MM$)	978	1622	4761	0	47948
Market Value (MM$)	978	1901	8361	0	154112
Inventory (M$)	942	136	495	0	6286
Business Segments	849	2.19	1.53	1	9
Loss	978	0.38	0.49	0	1
ROA	956	-0.10	0.52	-5.60	4.06
CATA	828	0.47	0.26	0.01	1
Inventory/Revenue	934	0.11	0.29	0	5.65
Quick Ratio	816	2.32	3.97	0.03	66
Goingconcern	978	0.05	0.22	0	1
Leverage	956	0.57	0.33	0.01	3.57
Dec	978	0.72	0.45	0	1
Industry dummies (NAICS code):					
Agriculture	978	0.00	0.05	0	1
Natural Recources	978	0.06	0.24	0	1
Utilities	978	0.05	0.21	0	1
Manufacturing/Construction	978	0.36	0.48	0	1
Trade (Retail, Wholesales)	978	0.07	0.26	0	1
Services	978	0.21	0.41	0	1
Information Technology	978	0.10	0.30	0	1
Financial Sector	978	0.14	0.35	0	1

Notes: MM$ is 1000 million, M$ is million dollars. Variables in Italics are dummies.

AUDITOR CHARACTERISTICS

	Arthur Andersen Clients				Rest.	Rest.	Ind. Exp.
	Fees, Act. ($)	Fees, Est. ($)	Asset (MM$)	*N*	(A.,AA)	(A.,GAO)	(MS)
PWC	751065	580513	2544	119	0.055	0.015	0.285
EY	783873	556035	1481	211	0.041	0.012	0.195
DT	850840	507440	2665	180	0.031	0.010	0.177
KPMG	798614	645523	1488	182	0.034	0.009	0.160
NonBig	127995	260171	62	74	0.005	0.005	0.005

Notes: First column is the average audit fee paid by former Andersen clients to their new auditor, the second is the estimated average fee generated by the regressions in Table 4. The third column is the average asset of, the fourth is the number of former Andersen clients by their new auditors. Restatement frequencies are calculated using two different databases: AuditAnalytics (Rest. (A.,AA)) and GAO and Who Audits America (Rest. (A.,GAO)). Industry Experience (Ind. Exp. (MS)) is calculated based on market shares.

Table 2: **Audit Fee regressions, all clients**

	(1)	(2)	(3)	(4)	(5)	(6)
Assets (log)	0.230***	0.207***	0.253***	0.328***	0.329***	0.256***
	(8.25)	(7.57)	(7.48)	(8.49)	(8.55)	(5.88)
Revenue (log)	0.157***	0.201***	0.166***	0.107***	0.102***	0.104***
	(6.70)	(8.57)	(6.68)	(3.59)	(3.43)	(3.51)
Revenue (zero)	2.380***	2.929***	2.536***	(dropped)	(dropped)	(dropped)
	(5.32)	(6.68)	(5.43)			
Market Value (log)	0.043**	0.060***	0.050***	0.052***	0.063***	0.068***
	(2.57)	(3.62)	(2.85)	(2.94)	(3.52)	(3.82)
Market Value (zero)	0.957***	1.277***	1.049***	1.060***	1.264***	1.368***
	(2.93)	(3.96)	(3.07)	(3.11)	(3.64)	(3.95)
Inventory (log)	0.070***	0.065***	0.073***	0.054***	0.047**	0.052***
	(3.73)	(3.59)	(3.74)	(2.66)	(2.31)	(2.59)
Inventory (zero)	1.159***	1.068***	1.210***	0.940***	0.823**	0.914***
	(3.75)	(3.57)	(3.80)	(2.89)	(2.53)	(2.82)
Business Segments	0.085***	0.082***	0.085***	0.082***	0.082***	0.082***
	(5.26)	(5.22)	(5.28)	(5.12)	(5.15)	(5.18)
Dec	0.123**	0.106**	0.118**	0.123**	0.119**	-0.350**
	(2.33)	(2.08)	(2.31)	(2.43)	(2.37)	(-2.45)
Loss		0.225***	0.171***	0.165***	0.153***	0.156***
		(4.33)	(3.17)	(3.07)	(2.86)	(2.93)
ROA		-0.195***	-0.188***	-0.161***	-0.127***	-0.137***
		(-4.40)	(-4.29)	(-3.64)	(-2.82)	(-3.05)
CATA			0.260**	0.525***	0.557***	0.557***
			(2.29)	(4.08)	(4.34)	(4.37)
Inventory/Revenue				0.173	0.251	0.222
				(0.92)	(1.33)	(1.18)
Quick Ratio				-0.043***	-0.036***	-0.036***
				(-4.83)	(-3.98)	(-4.01)
Goingconcern					0.163	0.150
					(1.45)	(1.35)
Leverage					0.196**	0.166**
					(2.32)	(1.98)
Assets (log) · *Dec*						0.089***
						(3.50)
Constant	-4.044***	-4.312***	-4.454***	-4.554***	-4.711***	-4.375***
	(-48.07)	(-47.62)	(-34.47)	(-34.08)	(-33.15)	(-25.68)
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	836	836	767	759	759	759
R^2	0.704	0.723	0.731	0.736	0.739	0.743
Adjusted R^2	0.700	0.719	0.727	0.731	0.734	0.738
F Statistics	196	179	158	148	131	126

*Notes: Dependent variable is the log of Audit Fees. t statistics are in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Variables in Italics are dummies.*

Table 3: **Audit Fee regressions, big 4 auditor clients only**

	(1)	(2)	(3)	(4)	(5)	(6)
Assets (log)	0.221***	0.196***	0.246***	0.332***	0.328***	0.237***
	(7.50)	(6.77)	(6.67)	(7.92)	(7.85)	(4.87)
Revenue (log)	0.159***	0.199***	0.167***	0.100***	0.096***	0.099***
	(6.42)	(7.96)	(6.35)	(3.20)	(3.05)	(3.18)
Revenue (zero)	2.573***	3.084***	2.555***	(dropped)	(dropped)	(dropped)
	(5.00)	(6.07)	(4.88)			
Market Value (log)	0.039**	0.057***	0.050**	0.052***	0.063***	0.072***
	(2.11)	(3.13)	(2.55)	(2.66)	(3.16)	(3.61)
Market Value (zero)	0.885**	1.249***	1.070***	1.088***	1.290***	1.469***
	(2.45)	(3.47)	(2.78)	(2.84)	(3.31)	(3.77)
Inventory (log)	0.099***	0.092***	0.097***	0.073***	0.065***	0.075***
	(4.94)	(4.68)	(4.56)	(3.12)	(2.80)	(3.22)
Inventory (zero)	1.693***	1.554***	1.656***	1.272***	1.160***	1.319***
	(5.03)	(4.73)	(4.66)	(3.40)	(3.09)	(3.52)
Business Segments	0.084***	0.082***	0.080***	0.077***	0.077***	0.076***
	(4.89)	(4.90)	(4.63)	(4.56)	(4.56)	(4.50)
Dec	0.127**	0.110**	0.114**	0.115**	0.109**	-0.461***
	(2.28)	(2.02)	(2.09)	(2.15)	(2.02)	(-2.72)
Loss		0.219***	0.177***	0.173***	0.162***	0.170***
		(4.02)	(3.12)	(3.06)	(2.85)	(3.01)
ROA		-0.161***	-0.146***	-0.122**	-0.097*	-0.107**
		(-3.25)	(-2.99)	(-2.47)	(-1.94)	(-2.15)
CATA			0.258**	0.547***	0.570***	0.565***
			(2.10)	(3.90)	(4.07)	(4.07)
Inventory/Revenue				0.196	0.318	0.249
				(0.67)	(1.08)	(0.85)
Quick Ratio				-0.044***	-0.038***	-0.038***
				(-4.78)	(-4.00)	(-4.07)
Goingconcern					0.078	0.104
					(0.55)	(0.74)
Leverage					0.221**	0.188*
					(2.19)	(1.88)
Assets (log) · *Dec*						0.103***
						(3.55)
Constant	-4.074***	-4.313***	-4.480***	-4.594***	-4.732***	-4.314***
	(-43.24)	(-42.84)	(-31.36)	(-30.43)	(-29.55)	(-21.82)
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	753	753	693	687	687	687
R^2	0.696	0.712	0.723	0.728	0.730	0.735
Adjusted R^2	0.692	0.707	0.718	0.722	0.724	0.729
F Statistics	170	152	136	128	113	109

*Notes: Dependent variable is the log of Audit Fees. t statistics are in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Variables in Italics are dummies.*

Table 4: **Audit Fee regressions for each auditor (pricing equations)**

	PWC	DT	EY	KPMG	NonBig
Assets (log)	0.072	0.297***	0.164	0.257**	0.005
	(0.67)	(3.79)	(1.28)	(2.40)	(0.04)
Revenue (log)	0.093	0.062	0.088	0.135*	0.169**
	(1.55)	(1.33)	(1.09)	(1.78)	(2.10)
Revenue (zero)	(dropped)	1.082	1.332	5.573***	2.619**
		(1.21)	(1.07)	(3.57)	(2.17)
Market Value (log)	0.093***	0.068*	0.077*	0.032	0.068
	(2.77)	(1.69)	(1.91)	(0.68)	(1.15)
Market Value (zero)	1.947***	1.323*	1.484*	0.843	1.118
	(3.00)	(1.70)	(1.84)	(0.93)	(1.12)
Business Segments	0.064*	0.123***	0.075**	0.022	0.023
	(1.90)	(4.29)	(2.21)	(0.50)	(0.45)
Dec	-0.827*	-0.358	-0.587	-0.694**	-0.559
	(-1.97)	(-1.22)	(-1.41)	(-2.04)	(-1.62)
CATA	0.671**	0.540**	0.470	0.475	-0.244
	(2.49)	(2.46)	(1.32)	(1.60)	(-0.64)
Leverage	0.314*	0.190	0.223	-0.011	0.187
	(1.70)	(1.26)	(0.98)	(-0.05)	(1.24)
Inventory (log)	0.215***	0.019	0.103**	0.077	-0.061
	(4.58)	(0.52)	(2.31)	(1.47)	(-1.05)
Inventory (zero)	3.476***	0.360	1.912**	1.211	-1.164
	(4.54)	(0.59)	(2.53)	(1.41)	(-1.37)
Quick Ratio	-0.028**	-0.022	-0.067**	-0.060***	-0.002
	(-2.10)	(-1.33)	(-2.17)	(-2.67)	(-0.08)
Loss	0.244*	0.068	0.252**	0.061	0.349**
	(1.97)	(0.71)	(2.02)	(0.48)	(2.11)
ROA	-0.210	-0.082	-0.015	-0.247*	-0.168*
	(-1.38)	(-1.42)	(-0.07)	(-1.66)	(-1.76)
Assets (log) · *Dec*	0.175**	0.096*	0.133*	0.136**	0.178*
	(2.35)	(1.97)	(1.93)	(2.20)	(1.80)
Constant	-4.015***	-4.310***	-3.970***	-3.941***	-3.200***
	(-8.60)	(-13.75)	(-8.23)	(-10.74)	(-6.65)
Industry Dummies	Yes	Yes	Yes	Yes	Yes
N	119	211	180	182	74
R^2	0.838	0.745	0.740	0.722	0.562
Adj. R^2	0.815	0.724	0.715	0.695	0.439
F-Statistics	36	35	29	27	5

*Notes: Dependent variable is the log of Audit Fees. t statistics are in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Variables in Italics are dummies.*

Table 5: **Multinomial logit, auditor characteristics, all clients**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rest. (A.I.)	-0.822				0.843			
	(-0.87)				(1.14)			
Rest. (A.I.T.)		-0.863				0.599		
		(-0.84)				(0.74)		
Rest. (A.,AA)			-2.525				7.408**	
			(-0.70)				(2.54)	
Rest. (A.,GAO)				-27.146*				22.283
				(-1.66)				(1.58)
Ind. Exp. (MS)	1.839***	1.821***	2.006***	2.326***				
	(5.53)	(5.56)	(4.02)	(4.84)				
Ind. Exp. (N)					0.001**	0.001***	0.000	0.001
					(2.56)	(2.62)	(0.70)	(1.03)
Audit Fees	0.233	0.232	0.242	0.247	0.288	0.293	0.230	0.268
	(1.31)	(1.31)	(1.34)	(1.36)	(1.51)	(1.53)	(1.26)	(1.43)
N	766	766	766	766	766	766	766	766
Pseudo R^2	0.016	0.016	0.016	0.017	0.007	0.007	0.010	0.008
χ^2	40	40	40	42	18	18	24	20
LogLikelihood	-1213	-1213	-1213	-1212	-1224	-1224	-1221	-1223

*Notes: Dependent variable is auditor choice. Non-parametric bootstrapped standard errors with 1000 replications. t statistics are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.*

Table 6: **Multinomial logit, auditor characteristics, big 4 auditor clients only**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rest. (A.I.)	-6.083***				-5.775***			
	(-3.75)				(-3.91)			
Rest. (A.I.T.)		-4.992***				-4.923***		
		(-3.32)				(-3.43)		
Rest. (A.,AA)			-23.544***				-24.688***	
			(-4.70)				(-4.34)	
Rest. (A.,GAO)				-85.983***				-101.430**
				(-4.48)				(-4.44)
Ind. Exp. (MS)	0.563	0.458	1.564***	1.533***				
	(1.33)	(1.10)	(3.41)	(3.28)				
Ind. Exp. (N)					0.000	0.000	0.002***	0.002***
					(0.18)	(0.09)	(2.61)	(2.96)
Audit Fees	-0.551**	-0.540**	-0.402**	-0.494**	-0.564**	-0.553**	-0.429**	-0.530**
	(-2.53)	(-2.51)	(-2.05)	(-2.37)	(-2.54)	(-2.53)	(-2.15)	(-2.49)
N	692	692	692	692	692	692	692	692
Pseudo R^2	0.012	0.010	0.016	0.015	0.011	0.009	0.015	0.015
χ^2	23	19	31	28	21	18	28	28
LogLikelihood	-948	-950	-944	-945	-949	-951	-945	-945

*Notes: Dependent variable is auditor choice. Non-parametric bootstrapped standard errors with 1000 replications. t statistics are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.*

Table 7: **Multinomial logit, auditor and client characteristics, all clients**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rest. (A.I.)	-4.253***	-4.769***	-4.843***	-5.530***	-5.422***	-5.283***	-5.354***
	(-2.70)	(-2.94)	(-2.97)	(-3.26)	(-3.22)	(-3.15)	(-3.20)
Rest. (A.I.) · *NonBig*							11.463
							(0.27)
Ind. Exp. (MS)	1.573***	1.520***	1.510***	1.253**	1.280**	1.265**	1.258**
	(3.28)	(3.07)	(3.03)	(2.38)	(2.40)	(2.34)	(2.33)
Ind. Exp. (MS) · *NonBig*							-19.155
							(-0.30)
Audit Fees	-0.620*	-0.595*	-0.577*	-0.537	-0.547	-0.541	-0.495
	(-1.91)	(-1.83)	(-1.68)	(-1.35)	(-1.37)	(-1.33)	(-1.22)
Audit Fees · *NonBig*							-10.180*
							(-1.71)
				PWC			
Assets	0.006	0.007	0.007	0.008	0.007	0.007	0.005
	(1.47)	(1.53)	(1.52)	(1.56)	(1.27)	(1.30)	(0.91)
Revenue	0.002	0.002	0.002	0.001	0.002	0.002	0.001
	(1.31)	(1.07)	(1.05)	(0.65)	(1.20)	(1.23)	(0.43)
Market Value	0.001	0.001	0.001	0.001	0.001	0.001	0.000
	(0.28)	(0.27)	(0.26)	(0.18)	(0.13)	(0.12)	(0.05)
Inventory	0.004	0.002	0.002	0.002	0.004	0.003	0.005
	(0.29)	(0.18)	(0.16)	(0.13)	(0.26)	(0.19)	(0.31)
Business Segments	-0.271**	-0.266**	-0.270**	-0.237*	-0.222*	-0.237*	-0.294*
	(-2.13)	(-2.10)	(-2.13)	(-1.77)	(-1.69)	(-1.80)	(-1.90)
Loss		-0.328	-0.230	-0.285	-0.249	-0.277	-0.686
		(-0.90)	(-0.59)	(-0.70)	(-0.59)	(-0.65)	(-1.32)
ROA			0.168	0.129	-0.138	-0.176	-0.194
			(0.59)	(0.43)	(-0.42)	(-0.53)	(-0.53)
CATA				1.668**	1.188	1.264	1.763*
				(2.05)	(1.23)	(1.30)	(1.66)
Inventory/Revenue				-0.756	-1.057	-1.178	-1.213
				(-0.53)	(-0.73)	(-0.81)	(-0.83)
Quick Ratio					0.098	0.057	0.084
					(0.82)	(0.47)	(0.64)
Goingconcern					-1.668	-1.487	-1.479
					(-0.50)	(-0.45)	(-0.46)
Leverage						-0.588	-0.812
						(-0.91)	(-1.14)
Constant	-0.778**	-0.467	-0.458	-1.306*	-1.155	-0.785	-1.549
	(-1.97)	(-0.94)	(-0.92)	(-1.67)	(-1.48)	(-0.87)	(-1.25)

Continued on the next page...

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				EY			
Assets	0.006	0.007	0.007	0.008	0.007	0.007	0.005
	(1.46)	(1.52)	(1.51)	(1.55)	(1.26)	(1.29)	(0.90)
Revenue	0.002	0.002	0.002	0.001	0.002	0.002	0.001
	(1.33)	(1.09)	(1.07)	(0.67)	(1.22)	(1.24)	(0.44)
Market Value	0.001	0.001	0.001	0.001	0.001	0.001	0.000
	(0.28)	(0.26)	(0.26)	(0.18)	(0.13)	(0.11)	(0.05)
Inventory	0.004	0.002	0.002	0.002	0.004	0.003	0.005
	(0.30)	(0.18)	(0.16)	(0.14)	(0.27)	(0.20)	(0.33)
Business Segments	-0.080	-0.086	-0.086	-0.046	-0.034	-0.050	-0.108
	(-0.72)	(-0.78)	(-0.78)	(-0.39)	(-0.29)	(-0.42)	(-0.77)
Loss		-0.738**	-0.736*	-0.834**	-0.786*	-0.818*	-1.226**
		(-2.13)	(-1.93)	(-2.09)	(-1.91)	(-1.95)	(-2.39)
ROA			-0.035	-0.066	-0.316	-0.350	-0.368
			(-0.12)	(-0.22)	(-0.95)	(-1.04)	(-0.98)
CATA				1.746**	1.297	1.400	1.911*
				(2.33)	(1.45)	(1.55)	(1.87)
Inventory/Revenue				-1.885	-2.171*	-2.279*	-2.353*
				(-1.54)	(-1.76)	(-1.86)	(-1.93)
Quick Ratio					0.094	0.056	0.083
					(0.78)	(0.47)	(0.63)
Goingconcern					-1.565	-1.416	-1.406
					(-1.31)	(-1.18)	(-1.28)
Leverage						-0.456	-0.674
						(-0.77)	(-1.03)
Constant	-0.444	0.056	0.067	-0.759	-0.607	-0.322	-1.091
	(-1.21)	(0.12)	(0.14)	(-1.01)	(-0.82)	(-0.38)	(-0.89)
				DT			
Assets	0.006	0.007	0.007	0.008	0.007	0.007	0.005
	(1.47)	(1.53)	(1.52)	(1.56)	(1.27)	(1.30)	(0.90)
Revenue	0.002	0.002	0.002	0.001	0.002	0.002	0.001
	(1.31)	(1.06)	(1.04)	(0.65)	(1.20)	(1.22)	(0.42)
Market Value	0.001	0.001	0.001	0.001	0.001	0.001	0.000
	(0.27)	(0.25)	(0.25)	(0.18)	(0.12)	(0.11)	(0.04)
Inventory	0.004	0.003	0.003	0.003	0.005	0.004	0.006
	(0.34)	(0.23)	(0.21)	(0.18)	(0.30)	(0.23)	(0.36)
Business Segments	-0.148	-0.152	-0.156	-0.127	-0.114	-0.131	-0.188
	(-1.27)	(-1.34)	(-1.36)	(-1.03)	(-0.94)	(-1.07)	(-1.29)
Loss		-0.727**	-0.623*	-0.672*	-0.682*	-0.722*	-1.137**
		(-2.11)	(-1.69)	(-1.74)	(-1.70)	(-1.77)	(-2.22)
ROA			0.189	0.240	0.078	0.046	0.030
			(0.69)	(0.81)	(0.25)	(0.15)	(0.09)
CATA				1.214	0.841	0.969	1.486
				(1.56)	(0.92)	(1.05)	(1.44)
Inventory/Revenue				-0.846	-1.084	-1.127	-1.169
				(-0.70)	(-0.89)	(-0.92)	(-0.97)
Quick Ratio					0.086	0.055	0.082
					(0.71)	(0.46)	(0.64)

Continued on the next page...

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				DT (continued)			
Goingconcern					-0.539	-0.472	-0.483
					(-0.93)	(-0.79)	(-0.75)
Leverage						-0.254	-0.465
						(-0.47)	(-0.76)
Constant	-0.519	-0.029	-0.023	-0.684	-0.576	-0.425	-1.204
	(-1.43)	(-0.06)	(-0.05)	(-0.90)	(-0.76)	(-0.50)	(-1.00)
				KPMG			
Assets	0.006	0.007	0.007	0.008	0.007	0.007	0.005
	(1.47)	(1.53)	(1.52)	(1.56)	(1.27)	(1.30)	(0.91)
Revenue	0.002	0.002	0.002	0.001	0.002	0.002	0.001
	(1.30)	(1.05)	(1.03)	(0.64)	(1.19)	(1.22)	(0.41)
Market Value	0.001	0.001	0.001	0.001	0.001	0.001	0.000
	(0.28)	(0.26)	(0.26)	(0.18)	(0.13)	(0.11)	(0.05)
Inventory	0.004	0.003	0.002	0.002	0.005	0.003	0.005
	(0.32)	(0.21)	(0.19)	(0.15)	(0.28)	(0.21)	(0.34)
Business Segments	-0.357***	-0.352***	-0.357***	-0.302**	-0.294**	-0.308**	-0.364**
	(-2.94)	(-2.95)	(-2.97)	(-2.35)	(-2.32)	(-2.42)	(-2.42)
Loss		-0.277	-0.167	-0.301	-0.237	-0.257	-0.662
		(-0.78)	(-0.44)	(-0.74)	(-0.57)	(-0.61)	(-1.28)
ROA			0.193	0.178	-0.076	-0.119	-0.135
			(0.76)	(0.64)	(-0.25)	(-0.38)	(-0.39)
CATA				2.380***	2.042**	2.081**	2.590**
				(3.00)	(2.17)	(2.18)	(2.47)
Inventory/Revenue				-0.716	-1.074	-1.279	-1.308
				(-0.64)	(-0.96)	(-1.15)	(-1.20)
Quick Ratio					0.071	0.018	0.045
					(0.60)	(0.16)	(0.36)
Goingconcern					-1.669	-1.374	-1.359
					(-1.14)	(-1.02)	(-0.98)
Leverage						-0.911	-1.130*
						(-1.48)	(-1.68)
Constant	-0.009	0.260	0.263	-1.029	-0.861	-0.288	-1.068
	(-0.03)	(0.55)	(0.55)	(-1.36)	(-1.15)	(-0.34)	(-0.87)
N	766	766	766	759	759	759	759
Pseudo R^2	0.108	0.112	0.113	0.121	0.128	0.130	0.132
χ^2	266	277	279	295	313	317	323
LogLikelihood	-1100	-1095	-1094	-1074	-1065	-1063	-1060

*Notes: Dependent variable is auditor choice. Non-parametric bootstrapped standard errors with 1000 replications. t statistics are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Variables in Italics are dummies. Base category is NonBig.*

Table 8: **Multinomial logit, auditor and client characteristics, big 4 auditor clients only**

	(1)	(2)	(3)	(4)	(5)	(6)
Rest. (A.I.)	-4.280***	-4.731***	-4.799***	-5.676***	-5.617***	-5.506***
	(-2.80)	(-3.01)	(-3.04)	(-3.36)	(-3.35)	(-3.30)
Ind. Exp. (MS)	1.580***	1.512***	1.513***	1.222**	1.231**	1.220**
	(3.13)	(2.95)	(2.93)	(2.32)	(2.32)	(2.28)
Audit Fees	-0.673**	-0.643*	-0.628*	-0.528	-0.540	-0.538
	(-2.01)	(-1.93)	(-1.76)	(-1.30)	(-1.32)	(-1.31)
				EY		
Assets	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.97)	(-0.87)	(-0.84)	(-0.78)	(-0.78)	(-0.79)
Revenue	0.000	0.000	0.000	0.000	0.000	0.000
	(0.36)	(0.32)	(0.31)	(0.21)	(0.21)	(0.20)
Market Value	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.35)	(-0.38)	(-0.38)	(-0.40)	(-0.39)	(-0.38)
Inventory	-0.000	-0.000	0.000	0.000	0.000	0.000
	(-0.00)	(-0.00)	(0.01)	(0.16)	(0.14)	(0.15)
Business Segments	0.194**	0.180**	0.185**	0.192**	0.191**	0.190**
	(2.17)	(2.03)	(2.06)	(2.10)	(2.07)	(2.06)
Loss		-0.400	-0.492*	-0.537*	-0.531*	-0.529*
		(-1.56)	(-1.72)	(-1.87)	(-1.82)	(-1.81)
ROA			-0.195	-0.179	-0.170	-0.163
			(-0.63)	(-0.56)	(-0.52)	(-0.49)
CATA				0.094	0.121	0.135
				(0.17)	(0.21)	(0.23)
Inventory/Revenue				-1.252	-1.233	-1.202
				(-0.91)	(-0.92)	(-0.89)
Quick Ratio					-0.004	-0.001
					(-0.08)	(-0.02)
Goingconcern					0.092	0.047
					(0.02)	(0.01)
Leverage						0.102
						(0.18)
Constant	0.334	0.522**	0.524**	0.544	0.542	0.478
	(1.55)	(2.07)	(2.07)	(1.44)	(1.43)	(0.92)
				DT		
Assets	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.12)	(-0.06)	(-0.06)	(-0.15)	(-0.12)	(-0.15)
Revenue	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.10)	(-0.14)	(-0.14)	(-0.12)	(-0.12)	(-0.12)
Market Value	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.69)	(-0.71)	(-0.70)	(-0.70)	(-0.69)	(-0.68)
Inventory	0.001	0.001	0.001	0.001	0.001	0.001
	(0.44)	(0.45)	(0.46)	(0.49)	(0.47)	(0.47)
Business Segments	0.123	0.112	0.112	0.109	0.108	0.106
	(1.33)	(1.21)	(1.21)	(1.16)	(1.13)	(1.11)

Continued on the next page...

	(1)	(2)	(3)	(4)	(5)	(6)
Loss		-0.392	-0.380	-0.379	-0.446	-0.450
		(-1.52)	(-1.36)	(-1.33)	(-1.50)	(-1.51)
ROA			0.031	0.118	0.201	0.211
			(0.10)	(0.37)	(0.56)	(0.58)
CATA				-0.426	-0.293	-0.257
				(-0.73)	(-0.48)	(-0.42)
Inventory/Revenue				-0.126	-0.078	0.031
				(-0.09)	(-0.05)	(0.02)
Quick Ratio					-0.012	-0.002
					(-0.25)	(-0.05)
Goingconcern					1.084	0.942
					(0.30)	(0.26)
Leverage						0.322
						(0.58)
Constant	0.261	0.438	0.436	0.607	0.557	0.349
	(1.11)	(1.64)	(1.63)	(1.52)	(1.36)	(0.63)
				KPMG		
Assets	0.000	0.000	0.000	0.000	0.000	0.000
	(0.07)	(0.04)	(0.04)	(0.12)	(0.13)	(0.15)
Revenue	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.19)	(-0.18)	(-0.19)	(-0.19)	(-0.20)	(-0.19)
Market Value	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.35)	(-0.32)	(-0.32)	(-0.39)	(-0.37)	(-0.38)
Inventory	0.000	0.000	0.000	0.000	0.000	0.000
	(0.25)	(0.25)	(0.26)	(0.27)	(0.26)	(0.27)
Business Segments	-0.088	-0.090	-0.090	-0.065	-0.072	-0.071
	(-0.90)	(-0.92)	(-0.92)	(-0.65)	(-0.72)	(-0.71)
Loss		0.045	0.060	-0.025	0.002	0.006
		(0.18)	(0.22)	(-0.09)	(0.01)	(0.02)
ROA			0.029	0.059	0.069	0.059
			(0.11)	(0.21)	(0.24)	(0.20)
CATA				0.703	0.840	0.805
				(1.22)	(1.37)	(1.31)
Inventory/Revenue				0.077	0.022	-0.088
				(0.06)	(0.02)	(-0.06)
Quick Ratio					-0.027	-0.039
					(-0.61)	(-0.76)
Goingconcern					0.024	0.185
					(0.01)	(0.05)
Leverage						-0.353
						(-0.64)
Constant	0.777***	0.738***	0.734***	0.277	0.291	0.513
	(3.37)	(2.77)	(2.74)	(0.68)	(0.71)	(0.95)
N	692	692	692	687	687	687
Pseudo R^2	0.043	0.047	0.048	0.052	0.055	0.056
χ^2	83	90	92	99	105	107
LogLikelihood	-918	-914	-913	-903	-900	-899

*Notes: Dependent variable is auditor choice. Non-parametric bootstrapped standard errors with 1000 replications. t statistics are in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Variables in Italics are dummies. Base category is PWC.*

Table 9: **Multinomial logit, auditor characteristics interacted with client characteristics, all clients**

	$X =$					
	Assets	Revenue	Market Value	Inventory	Business Segments	*Loss*
Rest. (A.I.)	-2.880	-4.048**	-4.021**	-5.343***	-3.870	-5.429***
	(-1.50)	(-2.09)	(-2.42)	(-2.70)	(-1.37)	(-2.86)
Rest. (A.I.) · X	-0.001	-0.001	-0.002	0.002	-0.492	1.952
	(-1.33)	(-0.77)	(-1.20)	(0.18)	(-0.64)	(0.47)
Ind. Exp. (MS)	1.279**	1.171*	1.011*	0.819	0.801	1.585**
	(2.14)	(1.88)	(1.71)	(1.30)	(0.87)	(2.47)
Ind. Exp. (MS) · X	-0.000	0.000	0.000	0.004	0.167	-1.169
	(-0.16)	(0.06)	(0.49)	(0.99)	(0.56)	(-1.08)
Audit Fees	-1.106**	-1.186**	-1.495***	-0.809*	-0.900	-0.642
	(-2.23)	(-2.34)	(-2.82)	(-1.70)	(-1.43)	(-1.22)
Audit Fees · X	0.000	0.000	0.000*	0.000	0.118	0.326
	(1.43)	(1.49)	(1.90)	(0.60)	(0.83)	(0.57)
				(···)		
N	759	759	759	759	759	759
Pseudo R^2	0.133	0.132	0.138	0.132	0.130	0.130
χ^2	325	322	336	322	318	318
LogLikelihood	-1059	-1060	-1054	-1060	-1063	-1062

	$X =$					
	ROA	CATA	Inventory/ Revenue	Quick Ratio	*Goingconcern*	Leverage
Rest. (A.I.)	-5.066***	-7.423**	-5.377**	-3.584*	-5.509***	-6.409
	(-2.98)	(-2.42)	(-2.46)	(-1.77)	(-3.20)	(-0.93)
Rest. (A.I.) · X	-5.364	10.258	1.734	-1.472	13.845	1.718
	(-0.63)	(1.23)	(0.06)	(-1.12)	(0.68)	(0.17)
Ind. Exp. (MS)	1.378**	2.416***	0.782	1.084	1.243**	1.552
	(2.50)	(2.58)	(1.16)	(1.60)	(2.26)	(1.20)
Ind. Exp. (MS) · X	2.008	-3.658*	6.935	0.128	-0.042	-0.489
	(1.39)	(-1.79)	(1.26)	(0.46)	(-0.01)	(-0.27)
Audit Fees	-0.538	-1.926***	-0.644	-0.753	-0.550	-0.088
	(-1.25)	(-2.67)	(-1.27)	(-1.46)	(-1.34)	(-0.09)
Audit Fees · X	-0.077	4.393***	1.120	0.227	0.848	-0.615
	(-0.08)	(2.82)	(0.24)	(0.87)	(0.38)	(-0.44)
				(···)		
N	759	759	759	759	759	759
Pseudo R^2	0.131	0.134	0.130	0.131	0.130	0.130
χ^2	319	328	318	319	318	317
LogLikelihood	-1062	-1057	-1062	-1062	-1063	-1063

*Notes: Dependent variable is auditor choice. The heading of a column indicates the variable which auditor characteristics are interacted with. All equations are estimated with the full set of client characteristics with base category NonBig, just as in model (6) of Table 7. Non-parametric bootstrapped standard errors with 1000 replications. t statistics are in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Variables in Italics are dummies.*

Table 10: Nested logit, auditor and client characteristics, all clients

	(1)	(2)	(3)	(4)	(5)	(6)
Rest. (A.I.)	-4.802***	-5.231***	-5.847***	-6.032***	-6.048***	-5.978***
	(-3.28)	(-3.45)	(-3.55)	(-3.67)	(-3.70)	(-3.67)
Rest. (A.I.) · *NonBig*	-11.853	-25.763	5.754	5.274	-1.826	4.159
	(-0.24)	(-0.51)	(0.12)	(0.10)	(-0.03)	(0.08)
Ind. Exp. (MS)	1.525***	1.479***	1.323***	1.186**	1.189**	1.193**
	(3.25)	(3.07)	(2.61)	(2.30)	(2.28)	(2.28)
Ind. Exp. (MS) · *NonBig*	-53.570	-70.214	-27.986	-21.606	-31.710	-20.004
	(-0.67)	(-0.85)	(-0.38)	(-0.27)	(-0.37)	(-0.24)
Audit Fees	-0.628**	-0.621**	-0.469	-0.437	-0.454	-0.446
	(-2.06)	(-2.06)	(-1.53)	(-1.25)	(-1.29)	(-1.27)
Audit Fees · *NonBig*	-0.705	-5.102	-9.267*	-9.943*	-10.711*	-10.590*
	(-0.19)	(-1.19)	(-1.82)	(-1.86)	(-1.85)	(-1.79)
			Upper Level:			
Assets	-0.005	-0.005	-0.006	-0.006	-0.005	-0.005
	(-1.09)	(-1.00)	(-1.08)	(-1.09)	(-0.87)	(-0.86)
Leverage	0.977**	1.086**	1.111**	1.163**	0.974*	0.736
	(2.41)	(2.58)	(2.57)	(2.51)	(1.84)	(1.27)
			Lower Level:			
			EY			
Revenue	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.12)	(-0.12)	(-0.14)	(-0.32)	(-0.30)	(-0.32)
Market Value	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-1.41)	(-1.48)	(-1.51)	(-1.55)	(-1.55)	(-1.54)
Inventory	0.000	0.000	0.000	0.000	0.000	0.000
	(0.24)	(0.17)	(0.38)	(0.63)	(0.59)	(0.61)
Business Segments	0.182**	0.170**	0.172**	0.183**	0.182**	0.182**
	(2.20)	(2.06)	(2.05)	(2.17)	(2.14)	(2.14)
Loss		-0.433*	-0.532**	-0.572**	-0.566**	-0.566**
		(-1.79)	(-2.01)	(-2.15)	(-2.11)	(-2.09)
ROA			-0.209	-0.187	-0.184	-0.177
			(-0.87)	(-0.78)	(-0.76)	(-0.73)
CATA			0.043	0.241	0.276	0.278
			(0.09)	(0.48)	(0.52)	(0.52)
Inventory/Revenue				-1.296	-1.301	-1.285
				(-1.06)	(-1.06)	(-1.05)
Quick Ratio					-0.006	-0.006
					(-0.17)	(-0.16)
Goingconcern						0.109
						(0.15)
Constant	0.323	0.527**	0.492	0.477	0.479	0.477
	(1.52)	(2.17)	(1.45)	(1.39)	(1.39)	(1.38)

Continued on the next page...

	(1)	(2)	(3)	(4)	(5)	(6)
				DT		
Revenue	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.42)	(-0.46)	(-0.55)	(-0.52)	(-0.53)	(-0.47)
Market Value	-0.000*	-0.000*	-0.000*	-0.000*	-0.000*	-0.000*
	(-1.71)	(-1.74)	(-1.74)	(-1.76)	(-1.75)	(-1.75)
Inventory	0.001	0.001	0.001	0.001	0.001	0.001
	(1.25)	(1.23)	(1.46)	(1.45)	(1.42)	(1.39)
Business Segments	0.124	0.112	0.103	0.111	0.106	0.110
	(1.49)	(1.34)	(1.22)	(1.30)	(1.23)	(1.28)
Loss		-0.399	-0.383	-0.393	-0.372	-0.458
		(-1.59)	(-1.39)	(-1.42)	(-1.33)	(-1.61)
ROA			0.015	0.112	0.116	0.196
			(0.05)	(0.38)	(0.39)	(0.65)
CATA			-0.387	-0.392	-0.289	-0.267
			(-0.80)	(-0.76)	(-0.53)	(-0.49)
Inventory/Revenue				-0.092	-0.139	-0.039
				(-0.08)	(-0.11)	(-0.03)
Quick Ratio					-0.022	-0.012
					(-0.53)	(-0.30)
Goingconcern						1.088
						(1.54)
Constant	0.245	0.430*	0.605*	0.578*	0.589*	0.532
	(1.11)	(1.73)	(1.76)	(1.66)	(1.69)	(1.52)
				KPMG		
Revenue	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.41)	(-0.40)	(-0.37)	(-0.35)	(-0.35)	(-0.36)
Market Value	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.80)	(-0.78)	(-0.89)	(-0.92)	(-0.89)	(-0.89)
Inventory	0.000	0.000	0.000	0.000	0.000	0.000
	(0.63)	(0.62)	(0.74)	(0.73)	(0.69)	(0.70)
Business Segments	-0.086	-0.088	-0.065	-0.060	-0.067	-0.067
	(-0.97)	(-1.00)	(-0.72)	(-0.66)	(-0.74)	(-0.74)
Loss		0.039	0.005	-0.020	0.006	0.007
		(0.16)	(0.02)	(-0.08)	(0.02)	(0.03)
ROA			0.048	0.059	0.065	0.070
			(0.19)	(0.23)	(0.25)	(0.27)
CATA			0.714	0.716	0.841	0.846
			(1.49)	(1.42)	(1.58)	(1.59)
Inventory/Revenue				0.107	0.059	0.055
				(0.10)	(0.05)	(0.05)
Quick Ratio					-0.027	-0.027
					(-0.76)	(-0.75)
Goingconcern						0.027
						(0.04)
Constant	0.754***	0.723***	0.291	0.247	0.263	0.262
	(3.42)	(2.83)	(0.79)	(0.66)	(0.71)	(0.70)

Continued on the next page...

	(1)	(2)	(3)	(4)	(5)	(6)
				NonBig		
Revenue	-0.002	-0.001	-0.001	-0.000	-0.001	-0.001
	(-1.05)	(-0.60)	(-0.34)	(-0.13)	(-0.27)	(-0.37)
Market Value	-0.001	-0.001	-0.000	-0.000	-0.000	-0.000
	(-1.09)	(-0.85)	(-0.37)	(-0.52)	(-0.49)	(-0.39)
Inventory	-0.004	-0.003	0.002	-0.003	-0.006	-0.005
	(-0.29)	(-0.21)	(0.19)	(-0.25)	(-0.45)	(-0.37)
Business Segments	0.112	0.213	0.231	0.288*	0.239	0.278*
	(0.64)	(1.42)	(1.51)	(1.66)	(1.38)	(1.67)
Loss		0.916*	1.022*	0.785	0.958	0.743
		(1.69)	(1.73)	(1.23)	(1.49)	(1.19)
ROA			0.064	0.026	0.034	0.187
			(0.28)	(0.12)	(0.16)	(0.83)
CATA			-1.984***	-2.090**	-1.908**	-1.774**
			(-2.68)	(-2.56)	(-2.10)	(-2.00)
Inventory/Revenue				1.137	1.214	1.263
				(1.39)	(1.39)	(1.46)
Quick Ratio					-0.095	-0.086
					(-1.17)	(-1.08)
Goingconcern						1.406*
						(1.79)
Constant	-3.656	-2.486	-0.433	0.885	0.583	1.149
	(-1.07)	(-0.78)	(-0.15)	(0.29)	(0.19)	(0.40)
Inclusive Value Big	-1.435	-0.674	-0.111	0.721	0.312	0.758
	(-0.69)	(-0.36)	(-0.06)	(0.37)	(0.15)	(0.40)
N	766	766	766	759	759	759
LogLikelihood	-1097	-1090	-1082	-1069	-1068	-1063

*Notes: Dependent variable is auditor choice. Non-parametric bootstrapped standard errors with 1000 replications. t statistics are in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Variables in Italics are dummies. Nesting structure: Big (PWC, EY, DT, KPMG), NonBig (NonBig). Base category is PWC at the lower and Big at the upper level. The model is estimated sequentially (LIML). The Inclusive Value of the degenerate nest (NonBig) is not identified; its value is 1. The LogLikelihood value is the sum of the loglikelihoods of the upper and lower levels.*

Table 11: **Multinomial logit with Industry Dummies, auditor and client characteristics, all clients**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rest. (A.I.)	-4.562**	-4.647**	-4.762**	-4.594**	-4.538**	-4.544**	-4.839**
	(-2.08)	(-2.15)	(-2.16)	(-2.06)	(-2.00)	(-1.97)	(-2.08)
Rest. (A.I.) · *NonBig*							55.609*
							(1.78)
Ind. Exp. (MS)	1.962***	1.893***	1.876**	2.089***	2.093***	2.067***	2.135***
	(2.76)	(2.58)	(2.51)	(2.77)	(2.70)	(2.58)	(2.64)
Ind. Exp. (MS) · *NonBig*							-30.606
							(-0.20)
Audit Fees	-0.231	-0.234	-0.206	-0.031	-0.013	-0.003	0.012
	(-0.89)	(-0.87)	(-0.73)	(-0.10)	(-0.04)	(-0.01)	(0.04)
Audit Fees · *NonBig*							-9.354
							(-1.40)
				PWC			
Assets	0.007	0.007	0.007	0.009	0.007	0.008	0.005
	(1.48)	(1.61)	(1.53)	(1.52)	(1.40)	(1.32)	(0.89)
Revenue	0.003	0.002	0.002	0.001	0.002	0.002	0.000
	(1.22)	(0.99)	(0.94)	(0.48)	(0.85)	(0.83)	(0.16)
Market Value	0.001	0.001	0.001	0.001	0.001	0.001	0.000
	(0.28)	(0.25)	(0.25)	(0.18)	(0.13)	(0.11)	(0.07)
Inventory	-0.005	-0.005	-0.005	-0.002	-0.000	-0.001	0.004
	(-0.43)	(-0.41)	(-0.41)	(-0.17)	(-0.03)	(-0.06)	(0.25)
Business Segments	-0.272**	-0.277**	-0.281**	-0.256*	-0.238*	-0.247*	-0.287*
	(-2.02)	(-2.03)	(-2.04)	(-1.74)	(-1.66)	(-1.69)	(-1.66)
Loss		-0.291	-0.213	-0.236	-0.184	-0.207	-0.613
		(-0.76)	(-0.51)	(-0.54)	(-0.40)	(-0.44)	(-1.06)
ROA			0.159	0.107	-0.136	-0.154	-0.173
			(0.52)	(0.33)	(-0.40)	(-0.44)	(-0.47)
CATA				1.639*	1.280	1.344	1.910*
				(1.78)	(1.20)	(1.25)	(1.65)
Inventory/Revenue				-0.463	-0.783	-0.840	-0.707
				(-0.33)	(-0.53)	(-0.56)	(-0.44)
Quick Ratio					0.087	0.058	0.064
					(0.73)	(0.44)	(0.48)
Goingconcern					-1.761	-1.647	-1.690
					(-0.49)	(-0.45)	(-0.45)
Leverage						-0.382	-0.560
						(-0.55)	(-0.74)
Constant	-0.930	-0.638	-0.623	-1.671	-1.491	-1.126	-2.167
	(-0.32)	(-0.21)	(-0.22)	(-0.60)	(-0.53)	(-0.36)	(-0.55)

Continued on the next page...

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				EY			
Assets	0.007	0.007	0.007	0.009	0.007	0.007	0.005
	(1.47)	(1.60)	(1.52)	(1.51)	(1.39)	(1.31)	(0.89)
Revenue	0.003	0.002	0.002	0.001	0.002	0.002	0.000
	(1.23)	(1.00)	(0.95)	(0.48)	(0.85)	(0.83)	(0.17)
Market Value	0.001	0.001	0.001	0.001	0.001	0.001	0.000
	(0.27)	(0.25)	(0.25)	(0.17)	(0.13)	(0.11)	(0.06)
Inventory	-0.005	-0.005	-0.005	-0.002	0.000	-0.000	0.005
	(-0.40)	(-0.39)	(-0.38)	(-0.13)	(0.01)	(-0.02)	(0.28)
Business Segments	-0.089	-0.100	-0.100	-0.067	-0.053	-0.062	-0.101
	(-0.71)	(-0.81)	(-0.78)	(-0.48)	(-0.38)	(-0.44)	(-0.59)
Loss		-0.574	-0.603	-0.621	-0.545	-0.570	-0.976*
		(-1.59)	(-1.56)	(-1.54)	(-1.27)	(-1.25)	(-1.78)
ROA			-0.085	-0.138	-0.370	-0.394	-0.415
			(-0.27)	(-0.41)	(-1.05)	(-1.02)	(-1.00)
CATA				2.047**	1.752*	1.812*	2.399**
				(2.48)	(1.79)	(1.79)	(2.20)
Inventory/Revenue				-2.099	-2.480*	-2.555*	-2.377
				(-1.51)	(-1.77)	(-1.85)	(-1.50)
Quick Ratio					0.074	0.045	0.050
					(0.62)	(0.34)	(0.38)
Goingconcern					-1.686*	-1.547**	-1.584
					(-1.70)	(-1.98)	(-1.44)
Leverage						-0.418	-0.593
						(-0.66)	(-0.83)
Constant	-0.392	-0.023	-0.039	-1.294	-1.129	-0.733	-1.803
	(-0.19)	(-0.01)	(-0.02)	(-0.58)	(-0.50)	(-0.28)	(-0.49)
				DT			
Assets	0.007	0.007	0.007	0.009	0.007	0.007	0.005
	(1.48)	(1.61)	(1.53)	(1.52)	(1.40)	(1.32)	(0.89)
Revenue	0.002	0.002	0.002	0.001	0.002	0.002	0.000
	(1.21)	(0.98)	(0.92)	(0.46)	(0.83)	(0.81)	(0.15)
Market Value	0.001	0.001	0.001	0.001	0.001	0.001	0.000
	(0.27)	(0.24)	(0.24)	(0.17)	(0.12)	(0.11)	(0.06)
Inventory	-0.004	-0.004	-0.004	-0.001	0.001	0.000	0.005
	(-0.36)	(-0.34)	(-0.34)	(-0.10)	(0.04)	(0.00)	(0.31)
Business Segments	-0.142	-0.153	-0.156	-0.133	-0.116	-0.126	-0.167
	(-1.10)	(-1.20)	(-1.18)	(-0.93)	(-0.83)	(-0.88)	(-0.98)
Loss		-0.677*	-0.634	-0.663	-0.671	-0.708	-1.113*
		(-1.78)	(-1.57)	(-1.56)	(-1.49)	(-1.49)	(-1.95)
ROA			0.079	0.083	-0.052	-0.063	-0.079
			(0.27)	(0.27)	(-0.16)	(-0.19)	(-0.23)
CATA				1.153	0.872	0.956	1.538
				(1.36)	(0.86)	(0.92)	(1.41)
Inventory/Revenue				-1.050	-1.297	-1.285	-1.128
				(-0.76)	(-0.96)	(-0.93)	(-0.72)
Quick Ratio					0.081	0.059	0.064
					(0.69)	(0.45)	(0.49)

Continued on the next page...

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				DT (continued)			
Goingconcern					-0.562	-0.515	-0.557
					(-0.88)	(-0.78)	(-0.77)
Leverage						-0.181	-0.358
						(-0.30)	(-0.51)
Constant	-0.710	-0.323	-0.328	-1.117	-1.004	-0.797	-1.863
	(-0.33)	(-0.14)	(-0.14)	(-0.49)	(-0.45)	(-0.31)	(-0.50)
				KPMG			
Assets	0.007	0.007	0.007	0.009	0.007	0.008	0.005
	(1.48)	(1.61)	(1.53)	(1.52)	(1.40)	(1.32)	(0.90)
Revenue	0.002	0.002	0.002	0.001	0.002	0.002	0.000
	(1.20)	(0.96)	(0.91)	(0.44)	(0.81)	(0.80)	(0.14)
Market Value	0.001	0.001	0.001	0.001	0.001	0.001	0.000
	(0.28)	(0.25)	(0.25)	(0.18)	(0.13)	(0.11)	(0.06)
Inventory	-0.005	-0.005	-0.005	-0.002	0.000	-0.000	0.005
	(-0.40)	(-0.38)	(-0.37)	(-0.12)	(0.01)	(-0.02)	(0.29)
Business Segments	-0.326**	-0.333**	-0.338**	-0.291**	-0.278**	-0.284**	-0.321*
	(-2.50)	(-2.55)	(-2.54)	(-1.99)	(-1.97)	(-1.97)	(-1.85)
Loss		-0.396	-0.317	-0.324	-0.255	-0.258	-0.661
		(-1.06)	(-0.79)	(-0.76)	(-0.55)	(-0.56)	(-1.16)
ROA			0.168	0.136	-0.095	-0.121	-0.140
			(0.56)	(0.42)	(-0.28)	(-0.33)	(-0.36)
CATA				2.239***	1.982**	1.984*	2.551**
				(2.64)	(2.02)	(1.92)	(2.33)
Inventory/Revenue				-1.233	-1.648	-1.830	-1.669
				(-0.90)	(-1.15)	(-1.28)	(-1.03)
Quick Ratio					0.069	0.029	0.034
					(0.59)	(0.22)	(0.26)
Goingconcern					-1.662	-1.410	-1.423
					(-1.00)	(-0.74)	(-0.77)
Leverage						-0.783	-0.964
						(-1.17)	(-1.29)
Constant	-2.556	-2.226	-2.212	-3.656	-3.498	-2.804	-3.865
	(-0.34)	(-0.29)	(-0.29)	(-0.48)	(-0.45)	(-0.35)	(-0.46)
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	766	766	766	759	759	759	759
Pseudo R^2	0.133	0.135	0.136	0.144	0.151	0.152	0.156
χ^2	328	334	336	352	369	372	380
LogLikelihood	-1069	-1066	-1065	-1046	-1037	-1035	-1031

*Notes: Dependent variable is auditor choice. All equations are estimated with the full set of industry dummies. Non-parametric bootstrapped standard errors with 1000 replications. t statistics are in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Variables in Italics are dummies. Base category is NonBig.*

Appendix Nested Logit normalisation with LIML

The usual procedure to estimate nested logit models is with full information maximum likelihood (FIML). FIML, as opposed to the limited information maximum likelihood, or sequential, estimation (LIML), has the distinct advantage that it is efficient, only one step and the standard errors are correct. However, the log likelihood function is not globally concave, which means that although there is a unique maximum, it can suffer from an abundance of local maxima. This can potentially complicate the numerical maximisation process to a great extent. Given the numerous variables in my model, the relatively small number of observations and the presence of a degenerate nest, FIML turned out to be difficult to estimate: it often failed to converge and when it did, it took 1500 or more iterations. Since the standard errors of the choice model have to be bootstrapped as a result of a generated regressor (Audit Fees), FIML estimation is simply not feasible in my case. Therefore, I estimated the nested logit models sequentially.

There is a delicate issue of normalizing the nested logit model in a sequential estimation procedure. As it has been pointed out by numerous studies (Koppelman and Wen 1998, Hunt 2000, Heiss 2002, Hensher and Greene 2002, Train 2003, Hensher et al 2005) some formulations (normalizations) are not consistent with the theory of random utility maximisation (RUM). Although, the issue has been addressed at length in a FIML framework, there has been little said about the normalization problem in sequential estimations.

The normalisation problem

Suppressing the index for the individual, let the utility from choosing alternative k in nest B be

$$U_k = W_k + \varepsilon_k = (V_B + v_B) + (V_{k|B} + v_{k|B})$$

where the first bracket is the utility associated with choosing an alternative from B and the second is the utility associated with kth alternative in B. Each utility component is broken down into an observed and unobserved (random) part. If $v_B = 0$, then errors within a nest do not share a common unobserved component and as a consequence there will be no correlation across the error terms within a nest. In nested logit models it is assumed that unobserved components at any given level and across levels are independently distributed, so formally

$$E(\varepsilon_{k\in B_j}\varepsilon_{l\in B_i}) = E(v_{B_j} + v_{k|B_j})(v_{B_i} + v_{l|B_i}) = E(v_{B_j}v_{B_i}) = \begin{cases} Var(v_{B_j}) \text{ if } j = i \\ 0 \text{ otherwise} \end{cases} \quad \text{for } k \neq l \qquad \text{(A.1)}$$

Therefore, the common random components can only induce correlation between two errors if the corresponding observations share the same nest. Given that the error terms are assumed to follow a type I extreme value distribution

$$Var(v_{k|B}) = \frac{\pi^2}{6(\lambda_B)^2}$$

That is, the variance of the error terms is inversely proportional to some scale parameter, λ_B (assumed to be common within a nest). Also,

$$Var(\varepsilon_{k\in B}) = Var(v_{k|B}) + Var(v_B) = \frac{\pi^2}{6(\lambda_b)^2}$$

It has been shown that (e.g. Hensher and Greene 2002)

$$\begin{aligned} V_B &= z'\gamma\lambda_b + \textstyle\sum_B \frac{\lambda_b}{\lambda_B} IV_B \text{ where} \\ IV_B &= \ln \textstyle\sum_k \exp(V_{k|B}\lambda_B) \end{aligned} \qquad \text{(A.2)}$$

where z is a vector of nest specific covariates with parameter vector γ. IV_B is the Inclusive Value (IV) or Logsum, the expected utility of choosing an alternative in nest B. In other words, it is a weighted average of the utilities associated with the alternatives within the nest, where the weights are probabilities of choosing each alternatives. The parameter of the inclusive value is the ratio between the two scale parameters, λ_B and λ_b. Two observations are immediate. First, since variances are positive, the scale parameters must be positive. Second, since the total variation contains both the upper and lower level variation it must be that $\lambda_b \leq \lambda_B$. This implies that the parameter of the inclusive values must lie between 0 and 1. If the value of the IV parameter is 1, this means $\lambda_b = \lambda_B$, which implies $Var(v_B) = 0$, that is, observations are not correlated within a nest (from (A.1)) and the nested logit specification is unnecessary, a simple multinomial logit is probably more appropriate. A value closer to 0 would imply strong correlation within nest.

λ_B and λ_b are not identified separately, only their ratio, therefore, a normalization is necessary. Naturally, there are two ways to normalize: either $\lambda_B = 1$ (RU1) or $\lambda_b = 1$ (RU2). Some algebraic manipulation shows that RU1 normalization is not consistent with RUM in general as it was pointed out in many articles (e.g. Koppelman and Wen 1998, Hunt 2000, Heiss 2002, Hensher and Greene 2002).

Normalisation with sequential estimation

Consider a RU2 specification. Then the RUM consistent sequential (LIML) estimation procedure is as follows. Suppose $V_{k|B} = x'\beta$. In the first stage estimation ("choice within nest"), we estimate a simple multinomial logit. Note that the variance of the error term is not identified at this stage, therefore - as a result of the error variance being normalized to $\pi^2/6$ - we actually estimate:

$$V_{k|B}\lambda_B = x'\beta\lambda_B \tag{A.3}$$

Therefore, without any explicit normalization, we can then calculate the inclusive values and estimate the model as in (A.2). In other words, the normalization is implicitly done in a sequential framework as it is clear from (A.3). Notice that this is quite different from a RUM consistent FIML estimation (RU2) where the "first stage" normalization must be explicitly carried out. In a sequential setting, therefore, we can obtain then $\hat{\beta}$ by multiplying the first stage estimates $\widehat{\beta\lambda_B}$ by $\widehat{1/\lambda_B}$ from the second stage. This is essentially the LIML equivalent of RU2 model.

Degenerate nests with sequential estimation

The presence of degenerate nest should not bring about any further complications in a sequential estimation in addition to the ones that emerge in FIML settings. In general, the IV parameter of the degenerate nest should be one but it's not identified if the model is RUM consistent since the IV parameter cancels out in the marginal probability. (e.g. Hunt 2000) If there is generic variable in the model, however, then the IV parameter will be identified even in RUM consistent models (RU2) but it will merely relax the restrictions imposed by the generic coefficient and as such will have little to do with the real IV parameter. (See, for instance, Heiss 2002)

In a sequential setting, therefore, as long as all coefficients are allowed to vary across partitions, then excluding the degenerate IV variable will deliver RUM consistent estimates. Note that in this case the degenerate IV variable is perfectly collinear with the variables of the degenerate nest in the second stage, therefore it cannot be included and/or constrained in any way. This is very intuitive and also algebraically obvious: since the inclusive value is basically a linear combination of the lower level variables, perfect collinearity will arise because the parameters of the lower level variables of the degenerate nest are estimated along with the IV variables in the second stage in a sequential framework. Observe that this is just another way of saying that the degenerate IV parameter plays no role whatsoever.

References

Allen, F. (1984): Reputation and Product Quality, RAND Journal of Economics, Vol. 15, No 3, 311-327

Allen, F.; Gale, D. (*2000*): *Comparing* Financial Systems, MIT Press

American Medical Association (2005): Competition in Health Insurance: a comprehensive study of US Markets

Amir, R.; Lambson, V. E. (2003): Entry, exit, and imperfect competition in the long run, Journal of Economic Theory, Elsevier, vol. 110(1), pages 191-203, May

Bar-Isaac, H. (2005): "Imperfect Competition and Reputational Commitment," Economics Letters, vol. 89, pp. 167-173, 2005.

Bar-Isaac, H. and Tadelis, S. (2008): "Seller Reputation", forthcoming in Kip Viscusi (ed): Foundations and Trends in Microeconomics, Now Publishers.

Bar-Yosef, S.; Sarath, B. (2005): Auditor Size, Market Segmentation and Litigation Patterns: a Theoretical Analysis, Review of Accounting Studies, 10, 59-92

Beck, T.; Demirgüç-Kunt, A.; Levine, R. (2006): Bank Concentration, Competition and Crises: First Results, Journal of Banking & Finance 30, 1581-1603

Beck, T.; Demirgüç-Kunt, A.; Levine, R. (2005): Bank Concentration and Fragility: Impact and Mechanics, NBER Working Paper No. 11500

Bikker, J. (2004): Competition and Efficiency in a Unified European Banking Market

Boyd, J.; Graham, S., L. (1996): Consolidation in US Banking: Implications for Efficiency and Risk, Federal Reserve Bank of Minneapolis, Working Paper 572

Boyd, J; Nicolo, G. (2005): The Theory of Bank Risk Taking and Competition Revisited, Journal of Finance, Vol. LX, No. 3, 1329-1343

Brewer III, E.; Jackson III, W., E. (2006): A Note on the "Risk Adjusted" Price-Concentration Relationship in Banking, Journal of Banking & Finance 30, 1041–1054

Caprio, G.; Klingebiel, D. (1996): Bank Insolvencies: Cross Country Experience, The World Bank, WP 1620

Chaney, P K; Philipich, K L (2002): Shredded Reputation: The Cost of Audit Failure, Journal of Accounting Research, v40, n4 (September 2002): 1221-45

Dana, J. D.; Fong, Y., F. (2008): Product Quality, Reputation, and Tacit Collusion, mimeo, Kellogg School of Management, 2008

Deshmukh, S. D.; Chikte S. D. (1976): Dynamic Pricing with Stochastic Entry, The Review of Economic Studies, Vol. 43, No. 1 (Feb., 1976), pp. 91-97

Diamond, D.W. (1989): "Reputation Acquisition in Debt Markets", Journal of Political Economy, Vol. 97, No. 4 (Aug., 1989), pp. 828-862

Economides, N. (1984): Symmetric Equilibrium Existence and Optimality in Differentiated Product Markets, Mimeo, Columbia University

Economist, The (2003): The next Arthur Andersen?, The Economist, 2/1/2003, Vol. 366 Issue 8309, p63-63, 2/5p

Eisenberg, T.; Macey, J. (2004): Was Arthur Andersen Different?: An Empirical Examination of Major Accounting Firms' Audits of Large Clients, 1 J. Empirical Legal Studies 263 2004

Ericson, R.; Pakes, A. (1995): Markov-Perfect Industry Dynamics: a Framework for Empirical Work, Review of Economic Studies 1995 62, 53-82

Farber, David B (2005): Restoring Trust after Fraud: Does Corporate Governance Matter?, Accounting Review, v80, n2 (April 2005): 539-61

Foreman, S. E.; Shea, D. G., (1999): 'Publication of Information and Market Response: The Case of Airline On-time Performance Reports,' Review of Industrial Organization, 14, pp. 147–162.

Francis, J.R. (2004): What do we know about audit quality?, The British Accounting Review 36 (2004) 345-368

Frazao, E.; Allshouse, J.E. (1996): Size and Growth of the Nutritionally Improved Foods Market, Agriculture Nutritionally Improved Foods Market. Agriculture Information Bulletin no, 723. Washington, D. C,: Economic Research Experiments."Service, U.S. Department of Agriculture

General Accounting Office (2008): Audits of Public Companies: Continued Concentration in Audit Market for Large Public Companies Does Not Call for Immediate Action, GAO-08-163

General Accounting Office (2006): Financial Restatements: Update of Public Company Trends, Market Impacts, and Regulatory Enforcement Activities, GAO-06-678

General Accounting Office (2003): Public Accounting Firms: Mandated Study on Consolidation and Competition, GAO-03-864

General Accounting Office (2002): Financial Statement Restatements: Trends, Market Impacts, Regulatory Responses, and Remaining Challenges, GAO-03-138

Glied, S. (2000): Managed Care, in Culyer, A.,J.; Newhouse, J.,P. (eds.): Handbook of Health Economics, Vol 1A. Amsterdam: Elsevier Science

Grimmett, G.; Stirzaker, D. (2001): Probability and Random Processes, Oxford University Press

Gul, F A; Lynn, S G; Tsui, J S L (2002): Audit Quality, Management Ownership, and the Informativeness of Accounting Earnings, Journal of Accounting, Auditing and Finance, v17, nl (Winter 2002): 25-49

Heiss, F. (2002): Structural choice analysis with nested logit models. Stata Journal, 2(3), 227–252

Hensher, D. A.; Greene, W. H. (2002): Specification and estimation of the nested logit model: alternative normalizations. Transportation Research 36(1): 1–17.

Hensher, D.A.; Rose, J.M., Greene, W.H. (2005): Applied choice analysis: A primer. Cambridge: Cambridge University Press

Hodgkin, D. (1996): Specialized Service Offerings and Patients' Choice of Hospital: The Case of Cardiac Catheterization, Journal of Health Economics, v15, n3 (June 1996): 305-32

Hopenhayn, H (1992): Entry, Exit, and firm Dynamics in Long Run Equilibrium, Econometrica, Vol. 60, No. 5 (Sep., 1992), pp. 1127-1150

Hörner, J. (2002): Reputation and Competition, American Economic Review Vol. 92. No 3, 644-663

Hunt, G. (2000): Alternative nested logit model structures and the special case of partial degeneracy. Journal of Regional Science 40: 89–113.

Jin, G. Z. (2005): Competition and Disclosure Incentives: an Empirical Studies of HMOs, RAND Journal of Economics Vol. 36, No. 1, 93-112

Jin, G Z; Leslie, P (2003): The effect of information on product quality: evidence from restaurant hygiene grade cards, Quarterly Journal of Economics, May 2003

Jiraporn, P. (2006): Shareholder Rights, Corporate Governance, and Arthur Andersen, Journal of Applied Finance, v16, n2 (Fall-Winter 2006): 12-23

Jovanovic, B (1982): Selection and the Evolution of Industry, Econometrica, Vol. 50, No. 3 (May, 1982), pp. 649-670

Kandori, M. (2006): "Repeated Games" The New Palgrave Dictionary of Economics, 2nd Edition, 2006

Kandori, M. (2002): Introduction to Repeated Games with Private Monitoring, Journal of Economic Theory, Vol. 102, Issue 1, January 2002, Pages 1-15

Keeley, M. (1990): Deposit Insurance, Risk and Market Power in Banking, American Economic Review 80, 1183-1200

Khurana, I K; Raman, K K (2004): Litigation Risk and the Financial Reporting Credibility of Big 4 versus Non-big 4 Audits: Evidence from Anglo-American Countries , Accounting Review, v79, n2 (April 2004): 473-95

Klein, B. and Leffler, K. B. (1981): "The Role of Market Forces in Assuring Contractual Performance," Journal of Political Economy, vol. 89, pp. 615-641, 1981.

Koppelman, F.S.; Bhat, C. (2006): A Self Instructing Course in Mode Choice Modelling: Multinomial and Nested Logit Models, prepared for U.S. Department of Transportation, Federal Transit Administration, available online on http://www.civil.northwestern.edu/people/koppelman.html.

Koppelman, F. S. and C. H. Wen (1998): Alternative nested logit models: structure, properties and estimation. Transportation Research-B 289–298.

Kranton, R. E. (2003): Competition and the Incentive to Produce High Quality, Economica, vol. 70, pp. 385-404, 2003

Krishnamurthy, S.; Zhou, J; Zhou, N (2006): Auditor Reputation, Auditor Independence, and the Stock-Market Impact of Andersen's Indictment on Its Client Firms, Contemporary Accounting Research/Recherche Comptable Contemporaine, v23, n2 (Summer 2006): 465-90

Loureiro ML, Gracia A, Nayga RM (2006): Do consumers value nutritional labels? European Review of Agricultural Economics, 2006 33(2):249-268

Louviere, J. J., Hensher D. A.; Swait J. D. (2000): Stated Choice Methods. Cambridge: Cambridge University Press.

Mathios, A. D., (2000): 'The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market,' Journal of Law and Economics, 43, pp. 651–677.

Maudos, J.; Guevara, F. (2004): Factors Explaining the Interest Margin in the Banking Sectors of the European Union, Journal of Banking & Finance 28, 2259-2281

Mocan, N. (2007): Can Consumers Detect Lemons? Information Asymmetry in the Market for Child Care, Journal of Population Economics 20:743–780, 2007

Murray, M P. (2006): The Bad, the Weak, and the Ugly: Avoiding the Pitfalls of Instrumental Variables Estimation, (October 2006). Available at SSRN: http://ssrn.com/abstract=843185

Palmrose, Z.; Scholz, S. (2004): The Circumstances and Legal Consequences of Non-GAAP Reporting: Evidence from Restatements, Contemporary Accounting Research v21, n1 (Spring 2004): 139-80

Pauly, M. (2004): Competition in Medical Services and the Quality of Care: Concepts and History, International Journal of Health Care Finance and Economics, 4, 113-130

Reed, B J; Trombley, M A; Dhaliwal, D S (2000): Demand for Audit Quality: The Case of Laventhol and Horwath's Auditees, Journal of Accounting, Auditing and Finance, v15, n2 (Spring 2000): 183-98

Reinstein, DA; Snyder CM (2005): The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics, Journal of Industrial Economics 03/2005, Vol 53/1, Pages: 27-51

Rivers, D.; Voung, Q. (1998) Limited information estimators and exogeneity tests for simultaneous probit models, Journal of Econometrics 39 (1988) 347-366

Rob, R.; Fishman, A. (2005): Is Bigger Better? Customer Base Expansion through Word-of-Mouth Reputation, Journal of Political Economy, Vol 113, No 5, 1146-1162

Romano, P.; Mutter, R. (2004): The Evolving Science of Quality Measurement for Hospitals: Implications for Studies of Competition and Consolidation, International Journal of Health Care Finance and Economics, 4, 131-157

Salop, S. (1979): Monopolistic Competition with Outside Goods, Bell Journal of Economics 10, 141-156

Shapiro, C. (1983): "Premiums for High Quality Products as Returns to Reputations," Quarterly Journal of Economics, vol. 98, pp. 659-680, 1983.

Sutton, J. (2006): Market Structure - Theory and Evidence, in Handbook of Industrial Organisation, Vol 3, edited by Robert Porter and Mark Armstrong.

Thomas, L.; Shane, S.; Weigelt, K. (1998): An Empirical Examination of Advertising as a Signal of Product Quality, Journal of Economic Behavior and Organization, v37, n4 (December 1998): 415-30

Tilis, L (2006): Audit quality and risk differences among auditors. ICFAI. Journal of Audit Practice, July, pp. 42 – 68. Also available at SSRN: http://ssrn.com/abstract=70024

Train, K. (2002): Discrete Choice Methods with Simulation. Cambridge University Press

Wong, H.; McNamara, P. (2004): Provider Competition and Health Care Quality: Challenges and Opportunities for Research, International Journal of Health Care Finance and Economics, 4, 99-11

Wooldridge, J. (2007): Control Function and Related Methods, What's New in Econometrics?, NBER Summer Institute 2007, http://www.nber.org/minicourse3.html